

Recovered Unified Process Views

Abram Hindle, Michael W. Godfrey, Richard C. Holt

Software Architecture Group
David R. Cheriton School of Computer Science
University of Waterloo

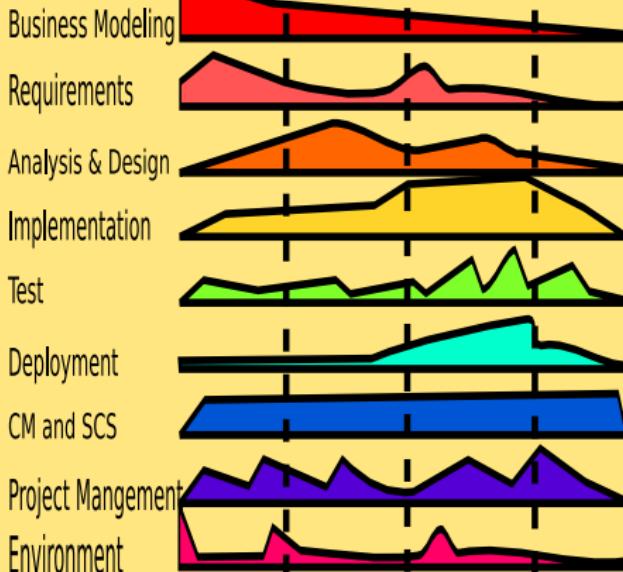
Canada

<http://swag.uwaterloo.ca/>

{ahindle,migod,holt}@cs.uwaterloo.ca

What are we going to do?

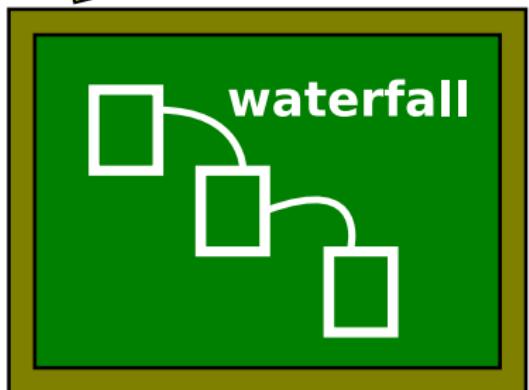
Theory



Practice

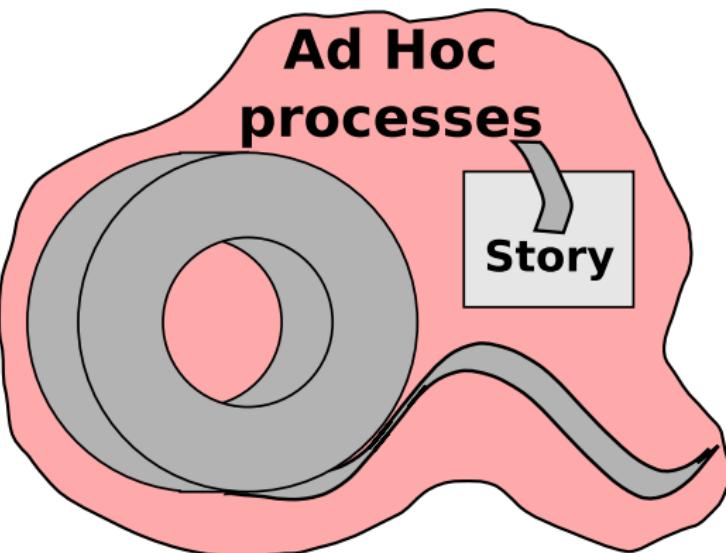


prescribed processes

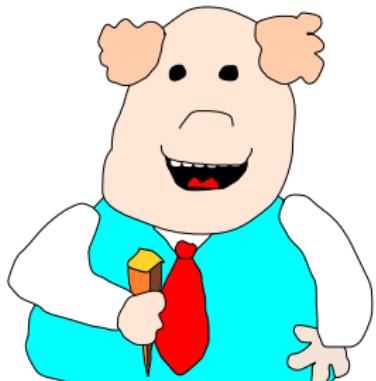


Formal Processes

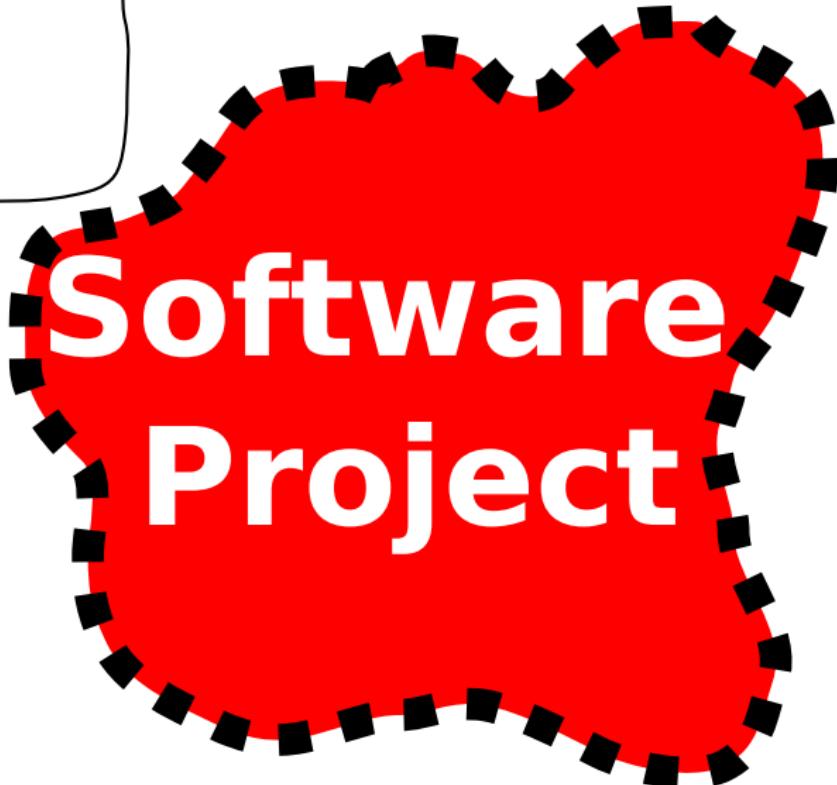
Process

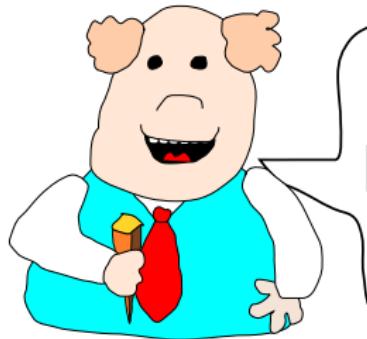


**what is going
on in this
project?**

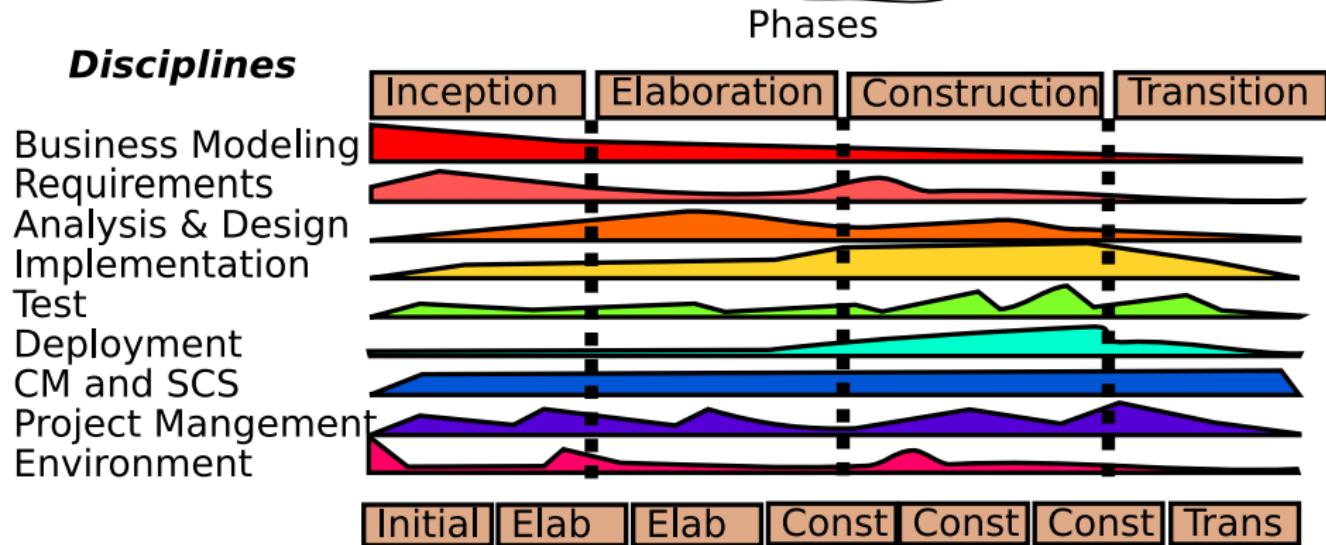


stakeholder



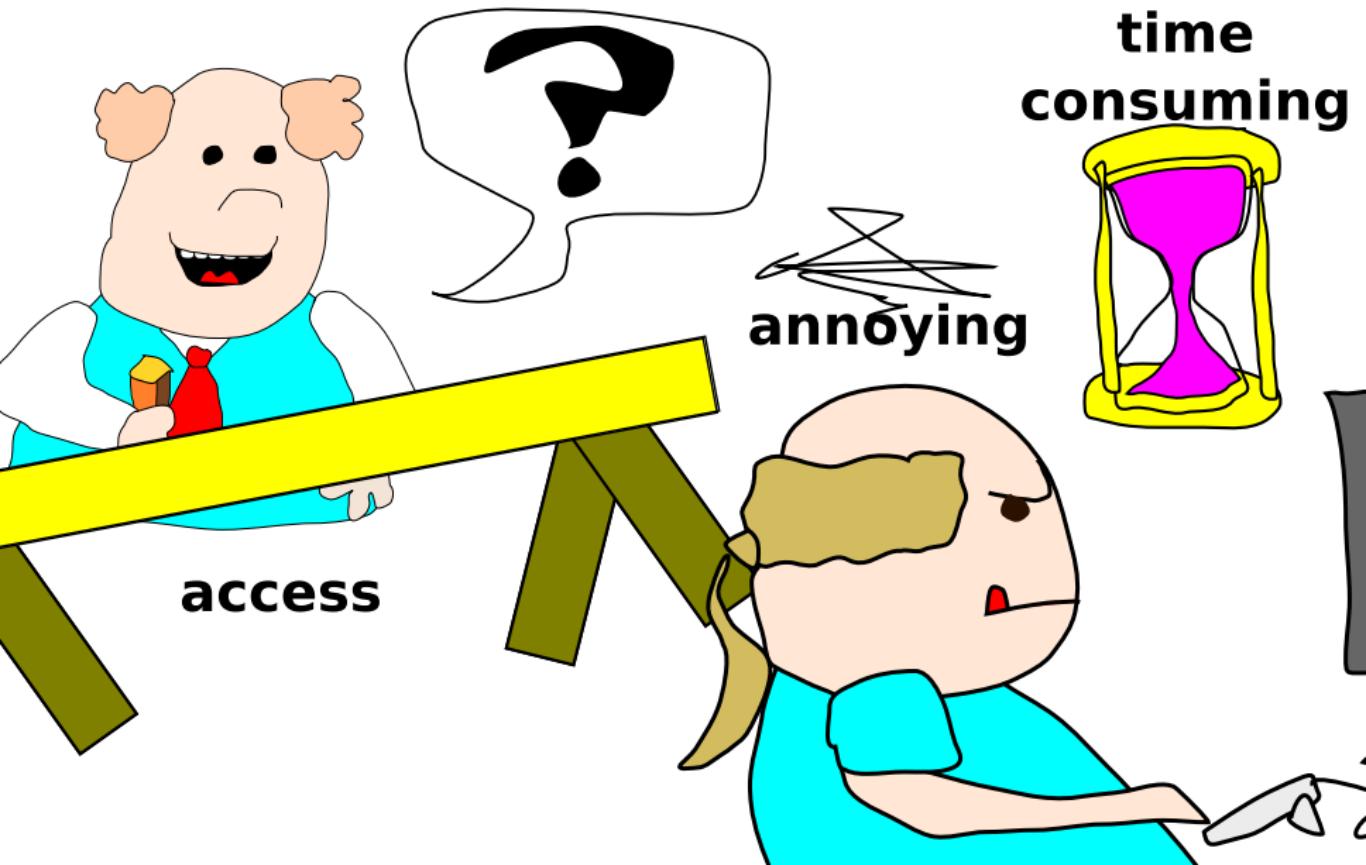


An overview of the project's processes and development would be nice!



Example UP Process

How to get an overview: Interviews



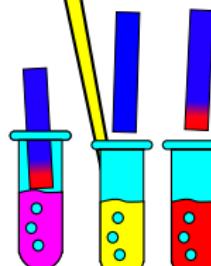
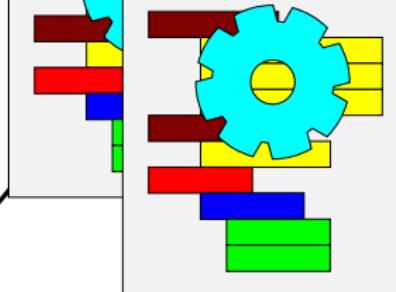
How to get an overview: Mining Software Repositories

Repository

stakeholder

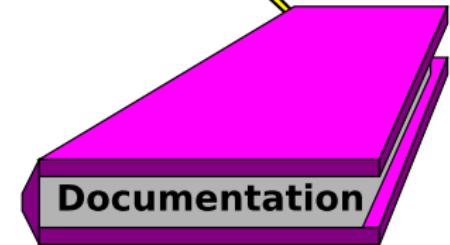
Source Code

Source Code

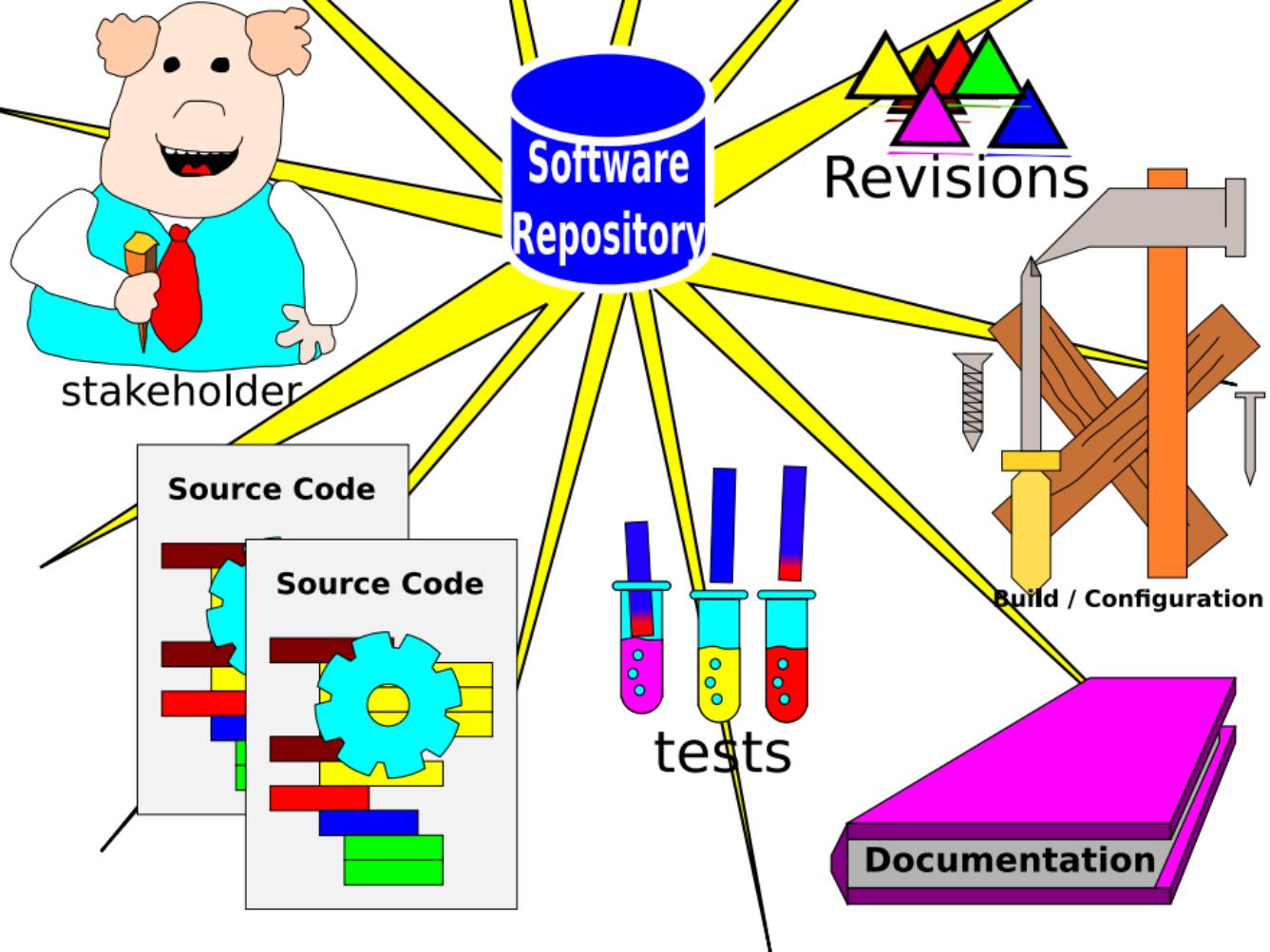


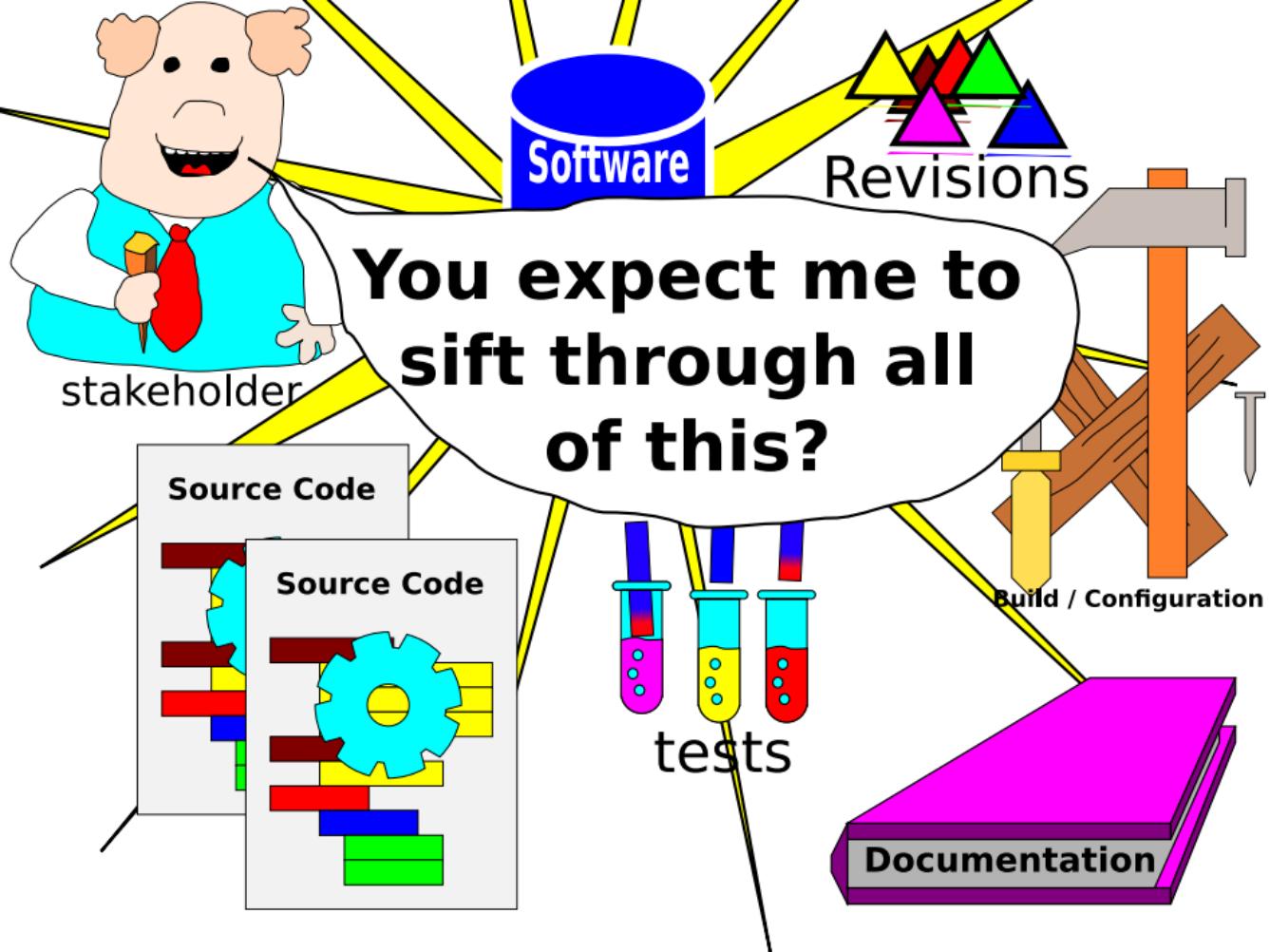
tests

Build / Configuration

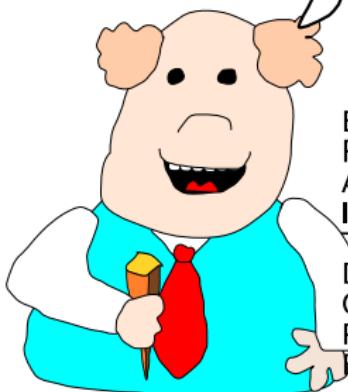
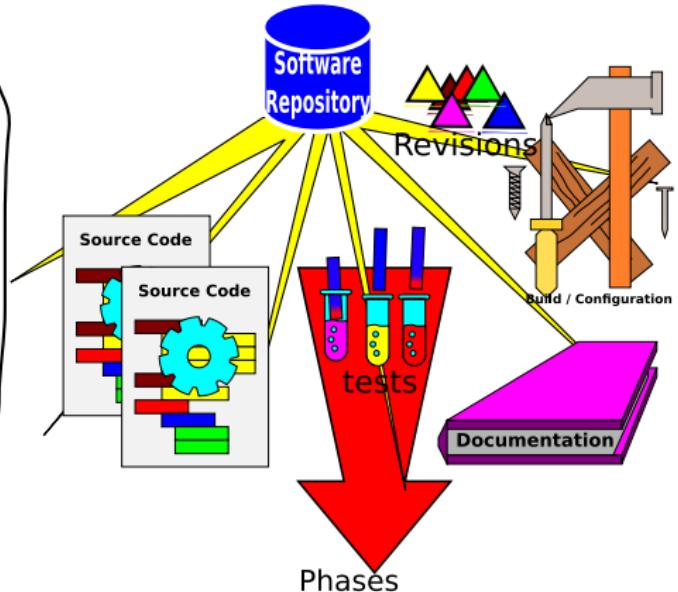


Documentation



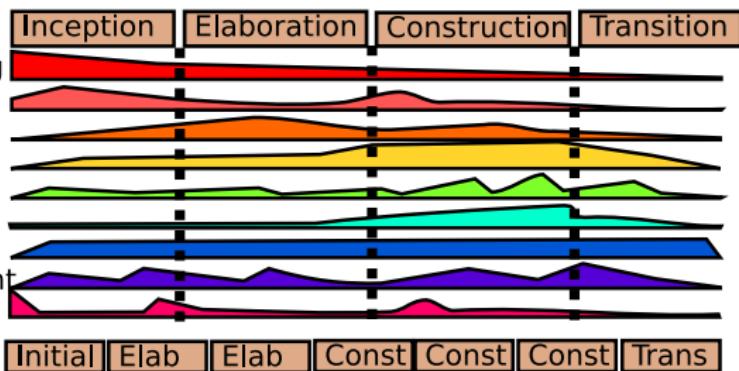


Can't we just summarize what is going on within this project?

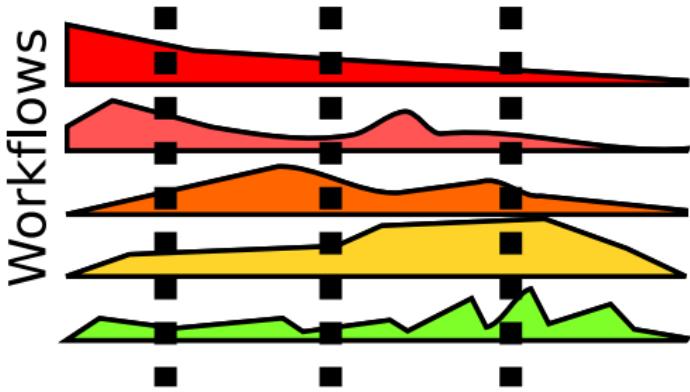


Disciplines

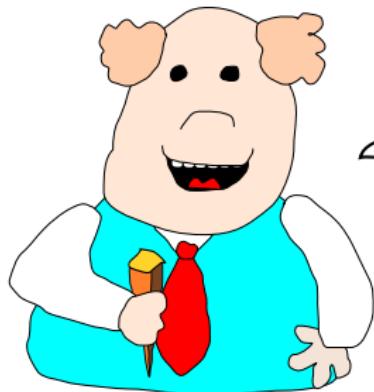
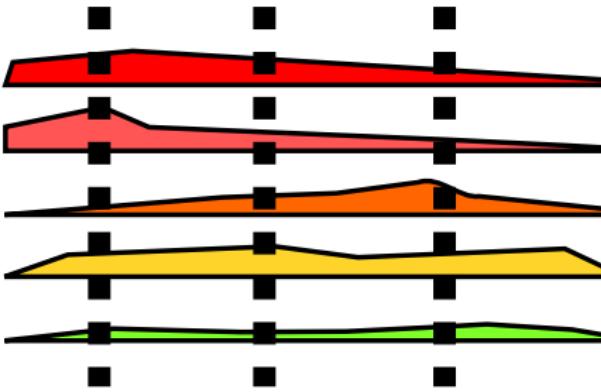
Business Modeling
Requirements
Analysis & Design
Implementation
Test
Deployment
CM and SCS
Project Management
Environment



Proposed Process



Recovered Process

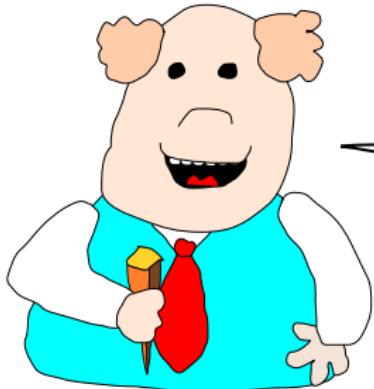
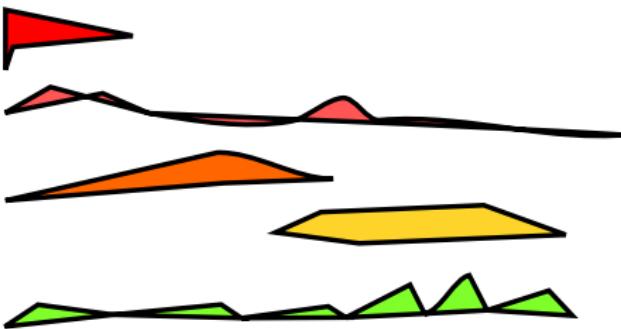


Is my proposed process actually being used?

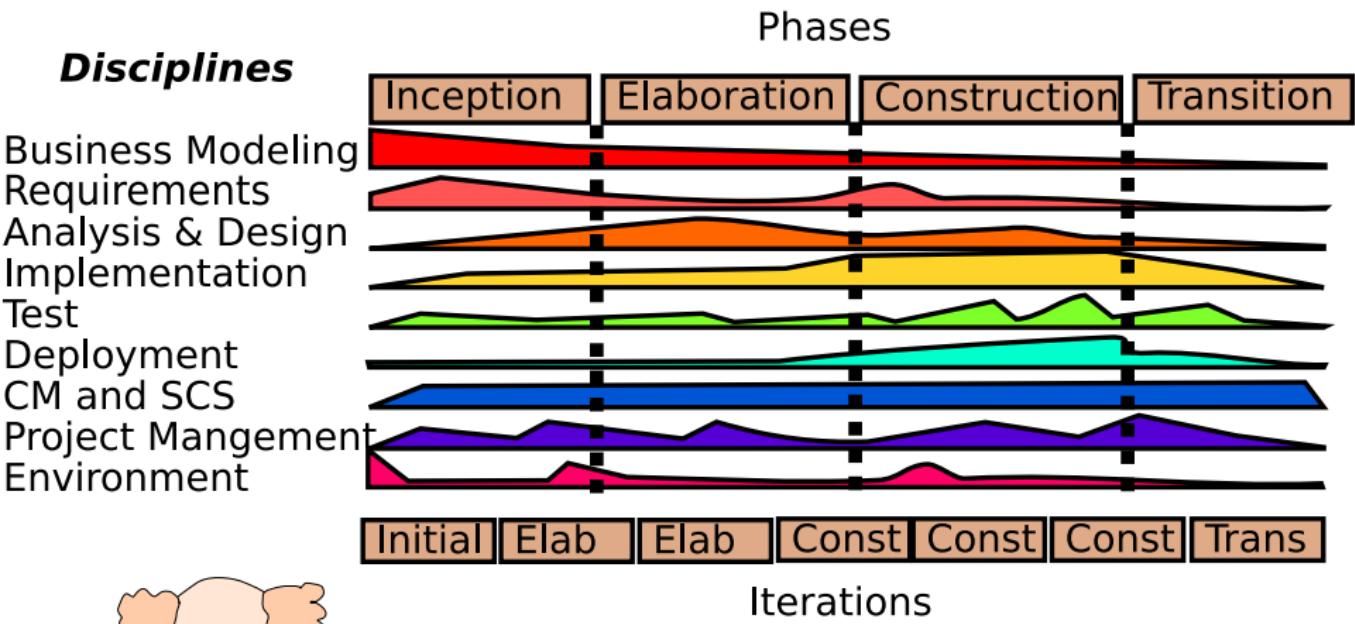
Proposed and Recovered Process Overlayed



Differences between Proposed and Recovered

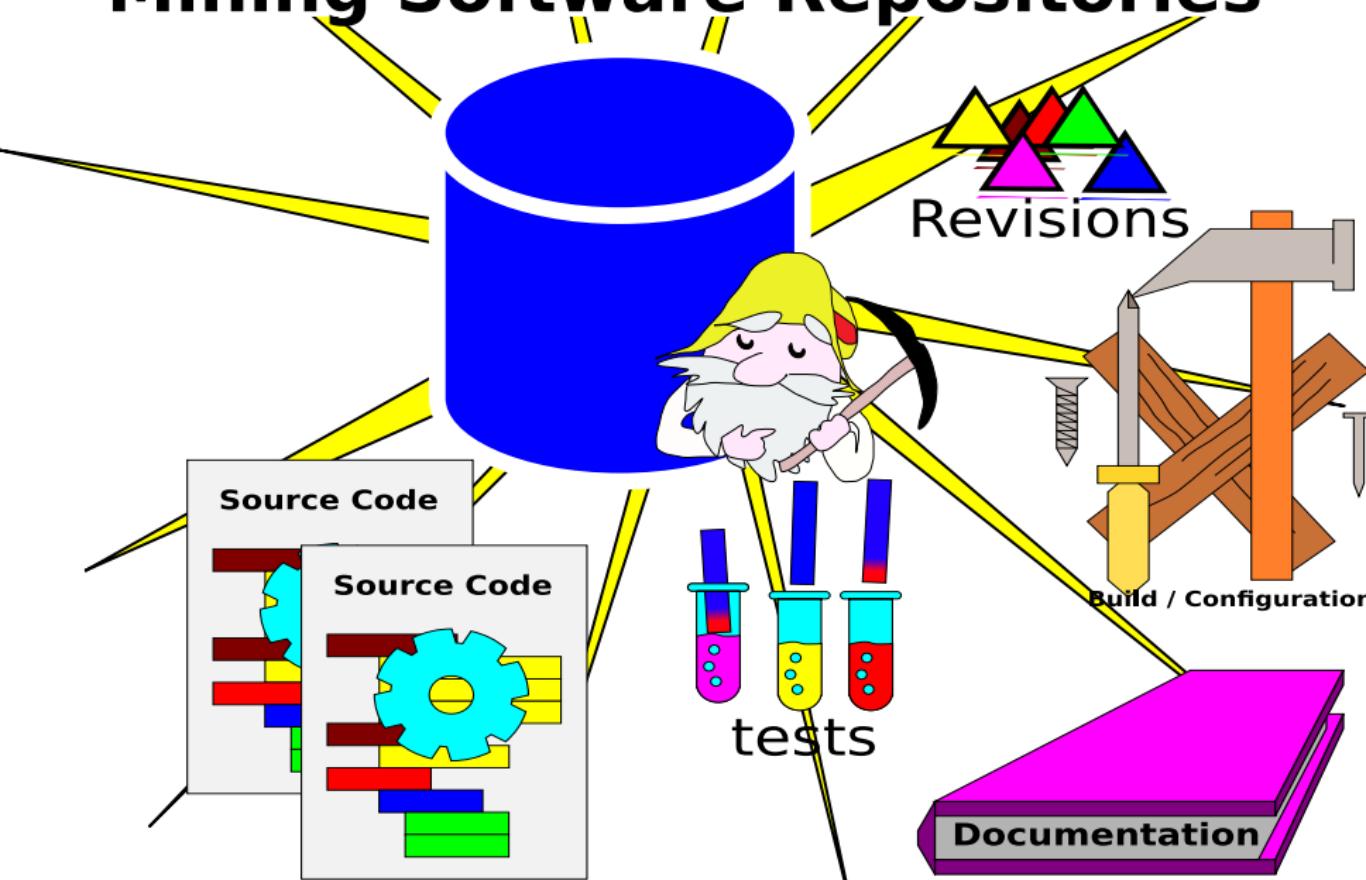


**I can compare
and contrast the
observed process
versus the
expected process!**

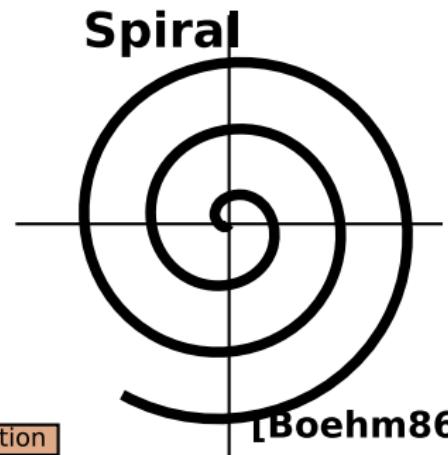
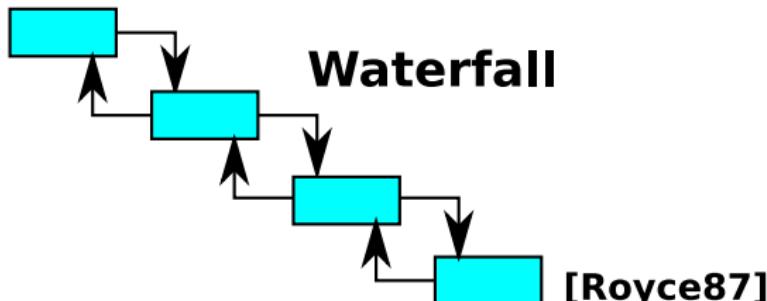


This **Unified Process** diagram shows different disciplines are used at different times.

Mining Software Repositories

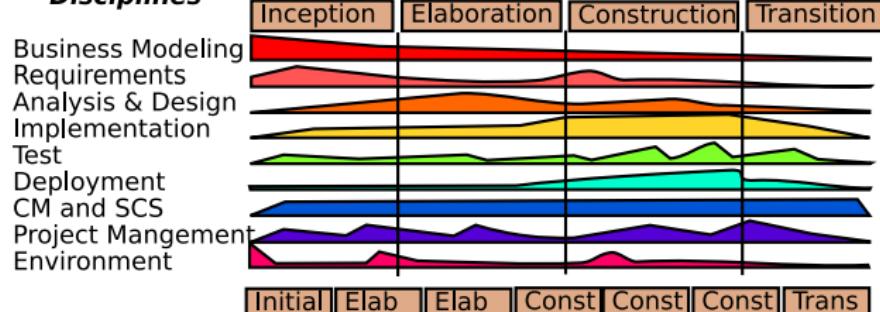


Software Development Processes



Unified Process

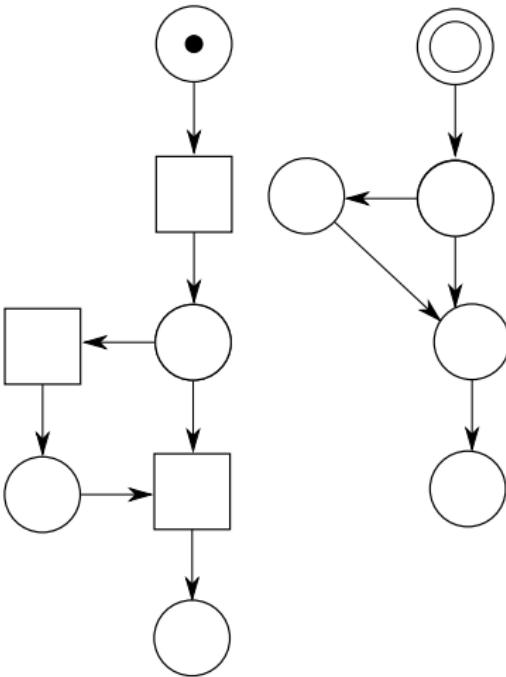
Disciplines



* CMM
* SDLC

[Jacobson99]

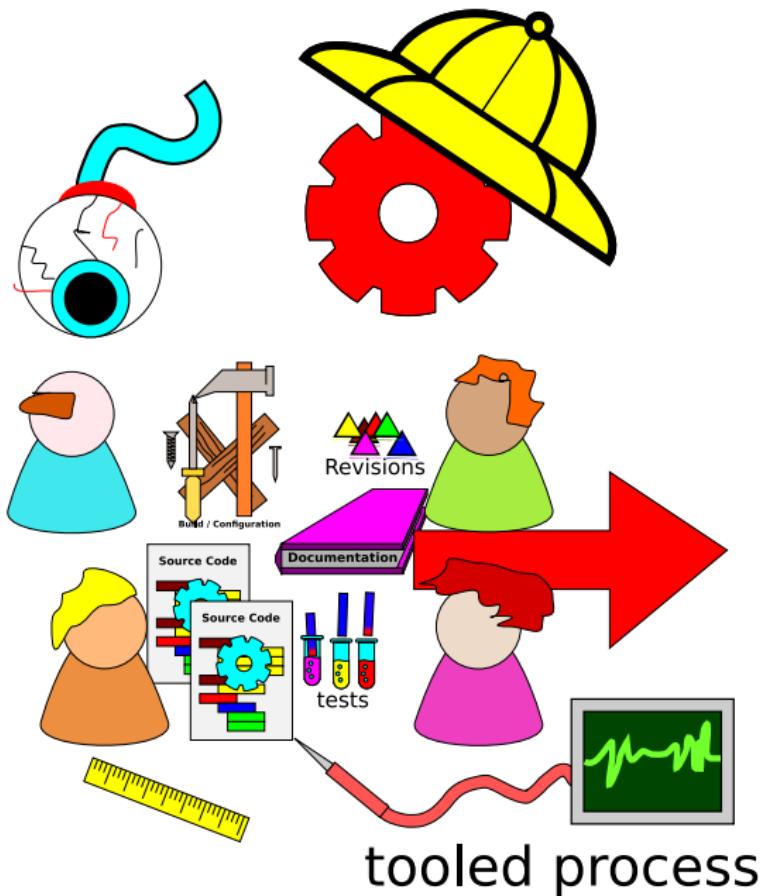
Process Mining



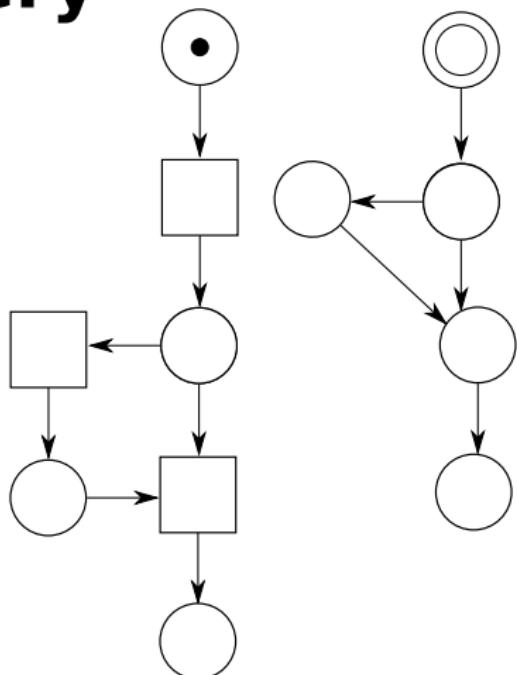
Petrinet FSM

[Aalst 2003]

Process Discovery

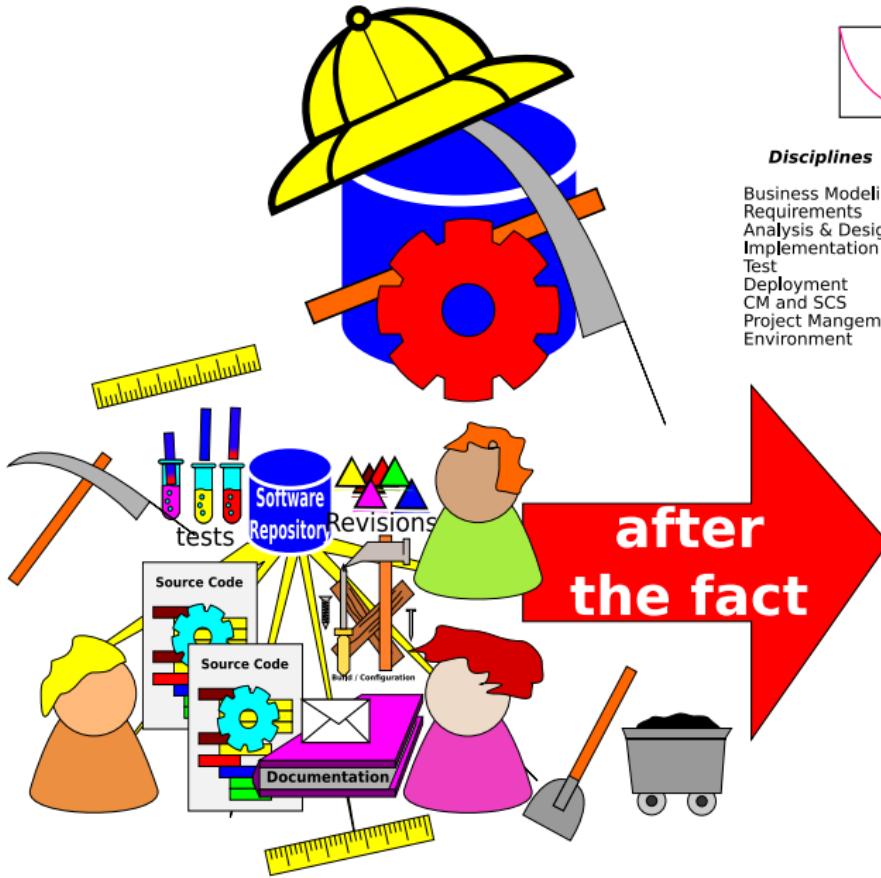


tooled process



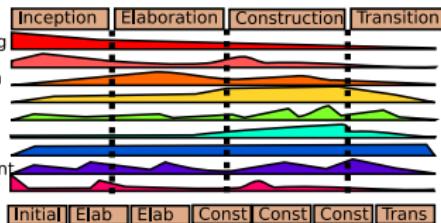
Petrinet FSM
[Cook96]

Process Recovery

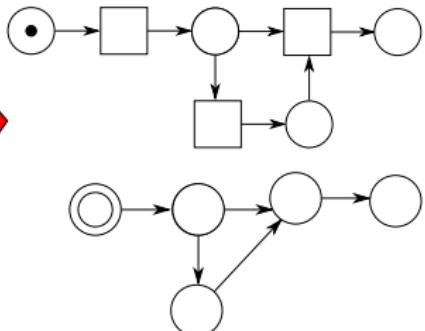


Disciplines

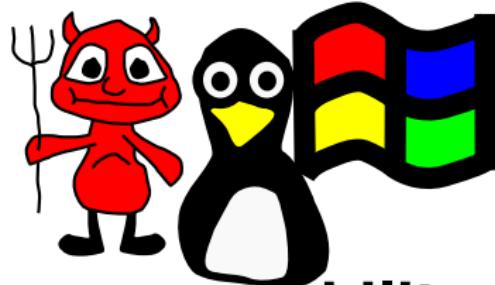
Business Modeling
Requirements
Analysis & Design
Implementation
Test
Deployment
CM and SCS
Project Management
Environment



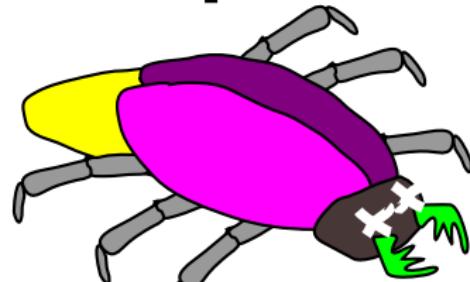
Initial Elab Elab Const Const Const Trans



Quality Related Non functional requirements



portability

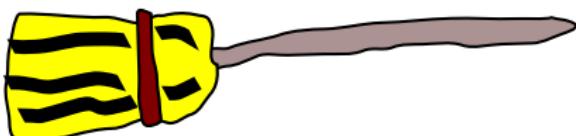


**reliability and
functionality**

(includes correctness)



usability

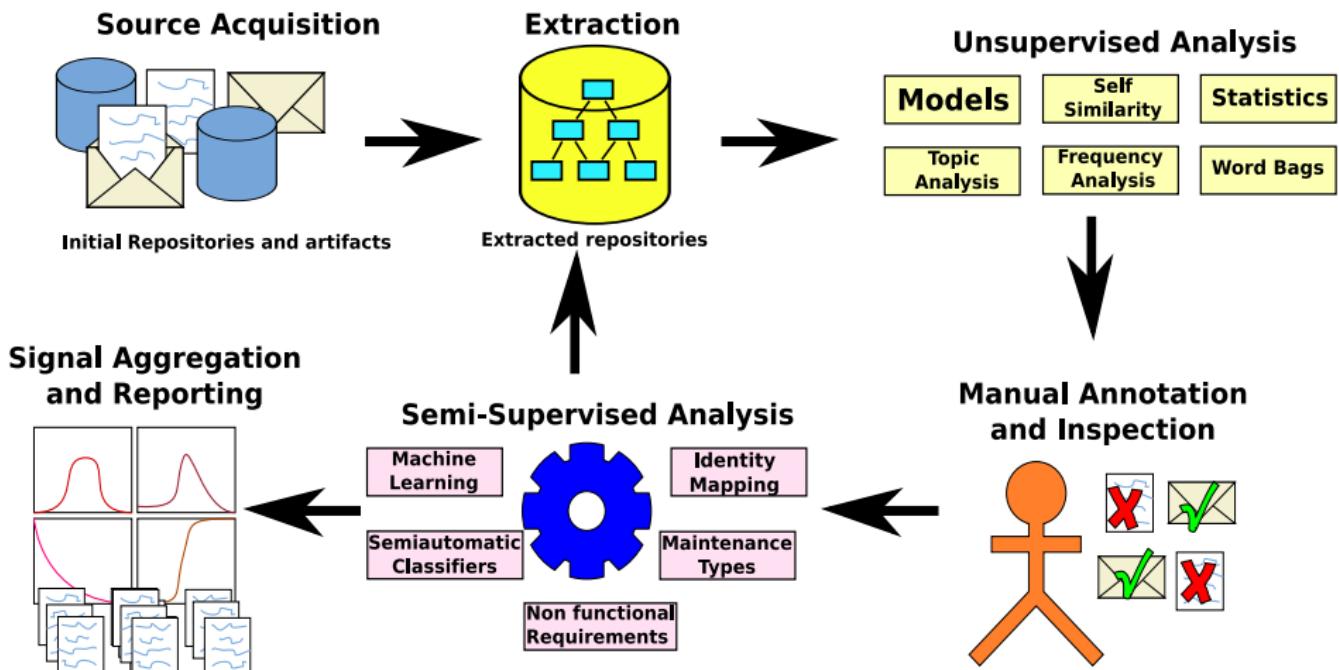


maintainability

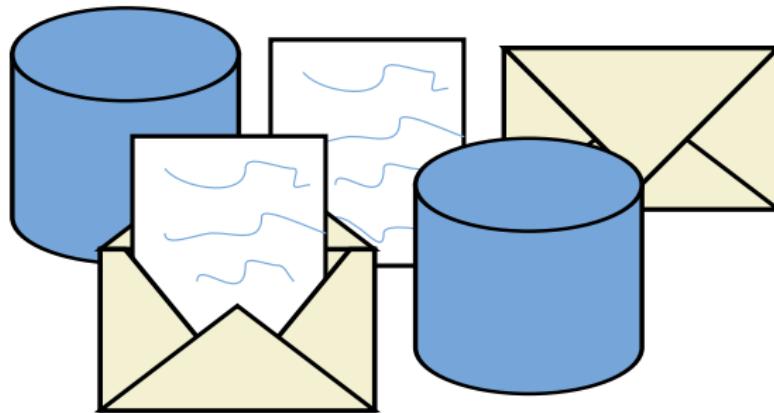


efficiency

Methodology: Recovered Unified Process Views

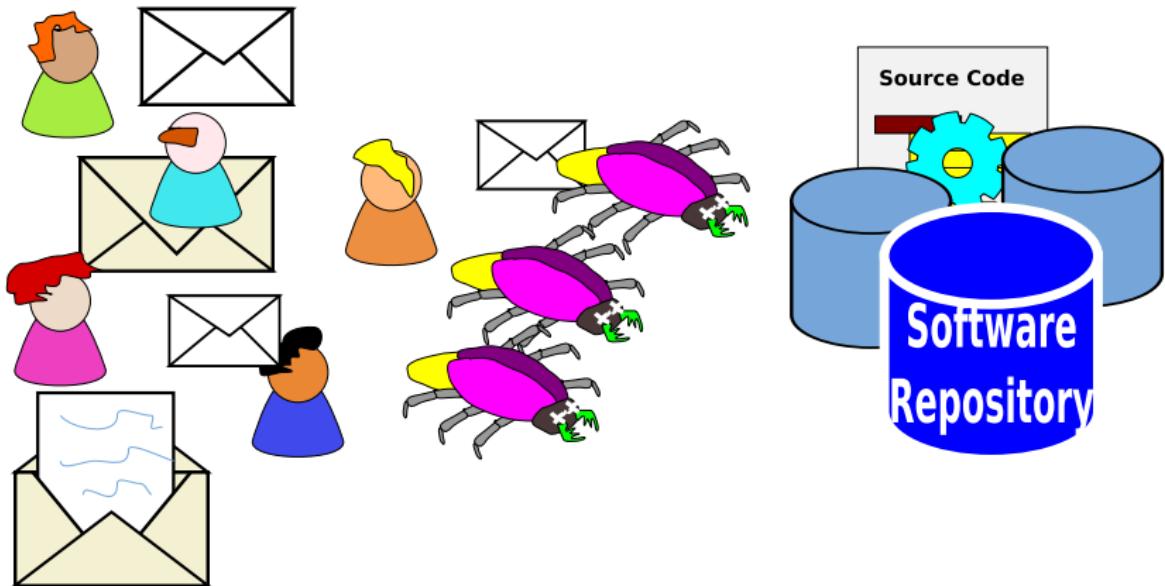


Source Acquisition



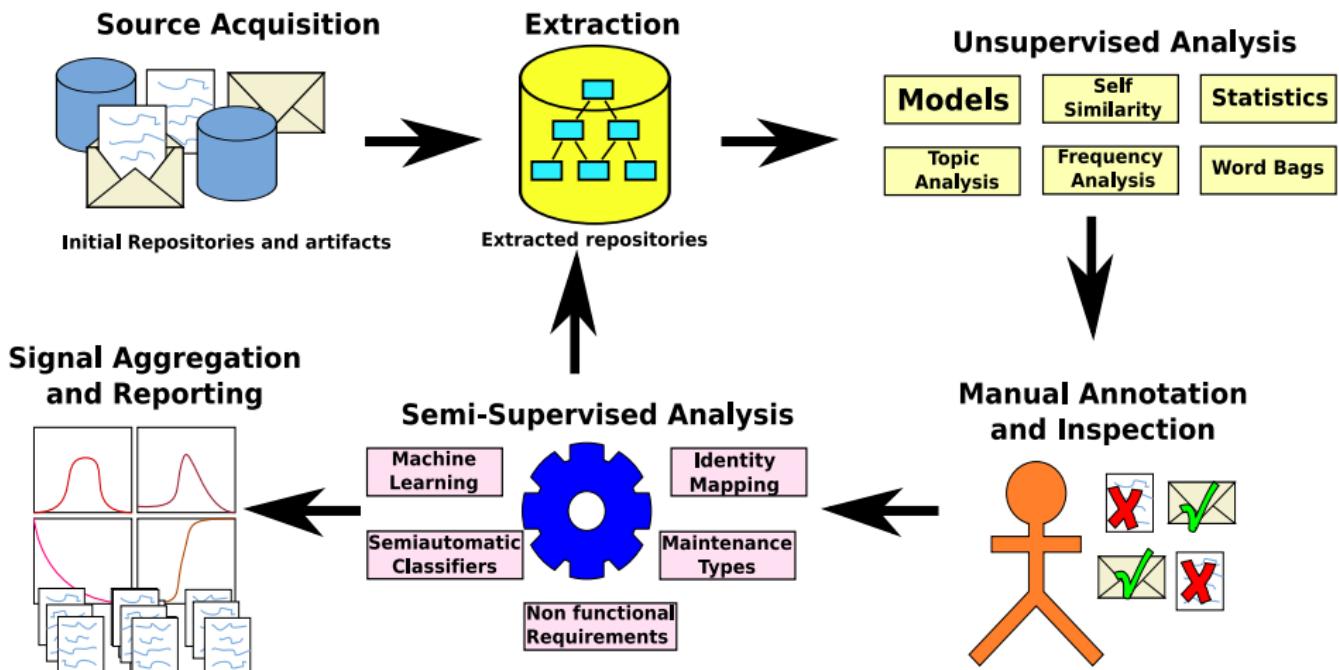
Initial Repositories and artifacts

Source Acquisition

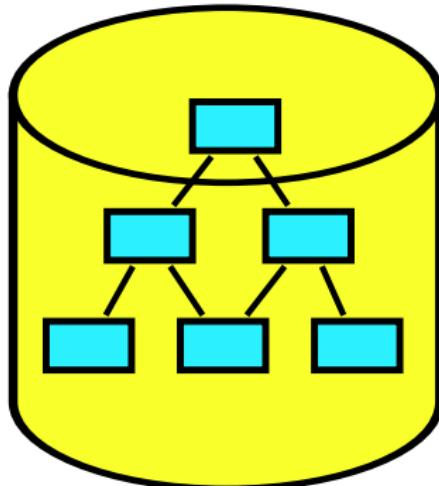


Initial Repositories and artifacts

Methodology: Recovered Unified Process Views

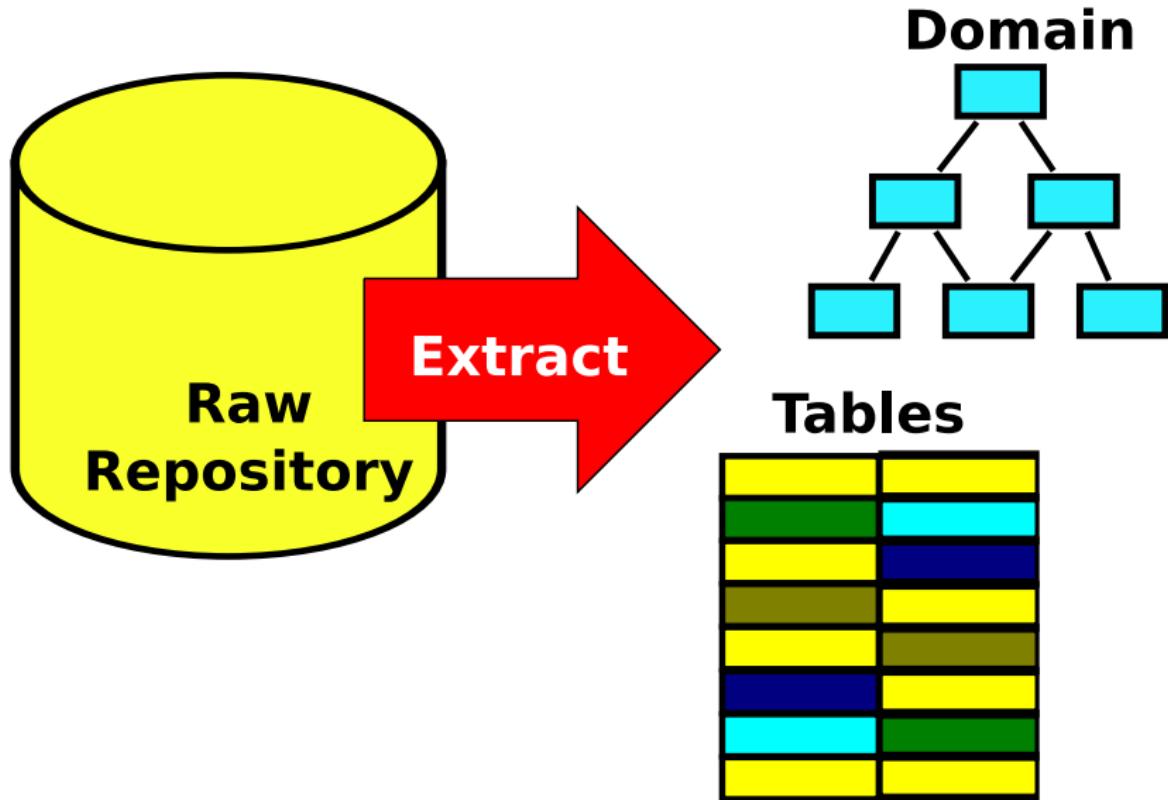


Extraction

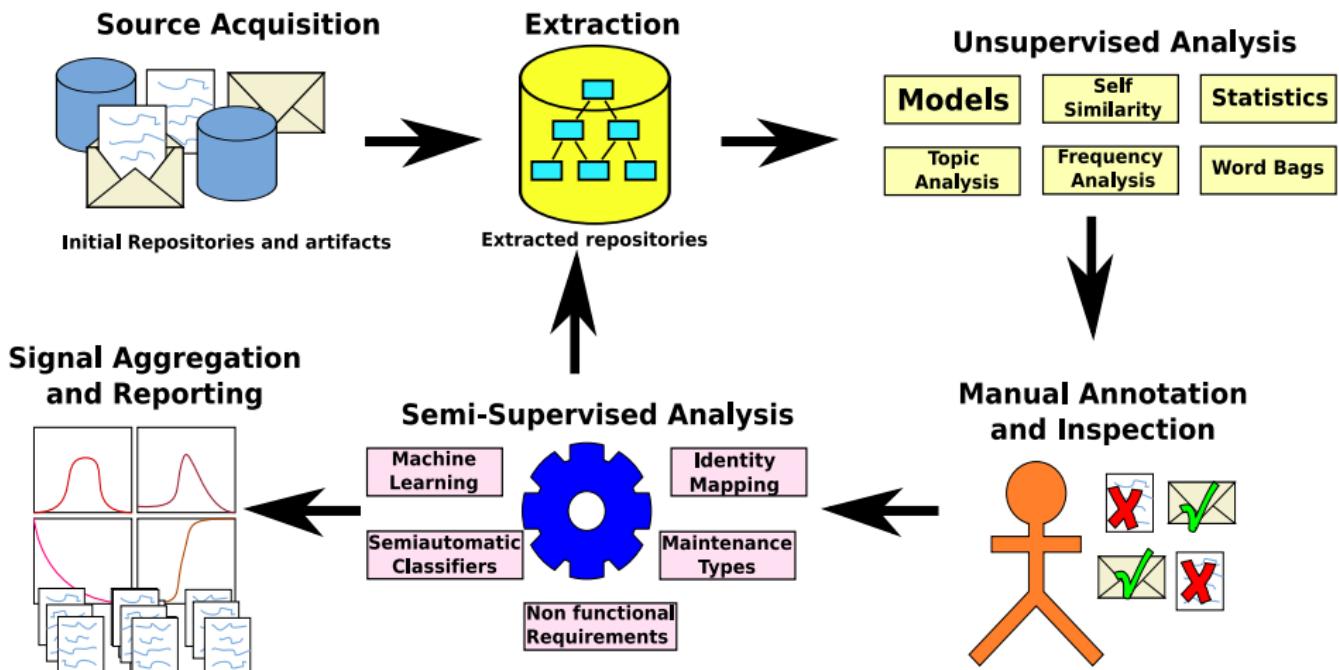


Extracted repositories

Extraction



Methodology: Recovered Unified Process Views



Unsupervised Analysis

Models

**Self
Similarity**

Statistics

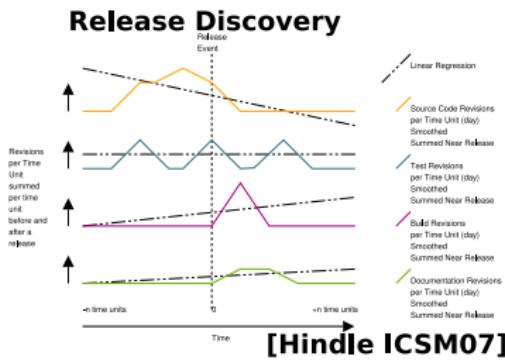
**Topic
Analysis**

**Frequency
Analysis**

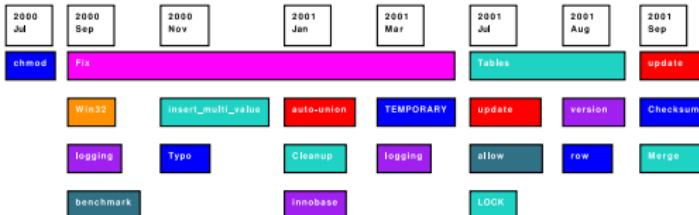
Word Bags



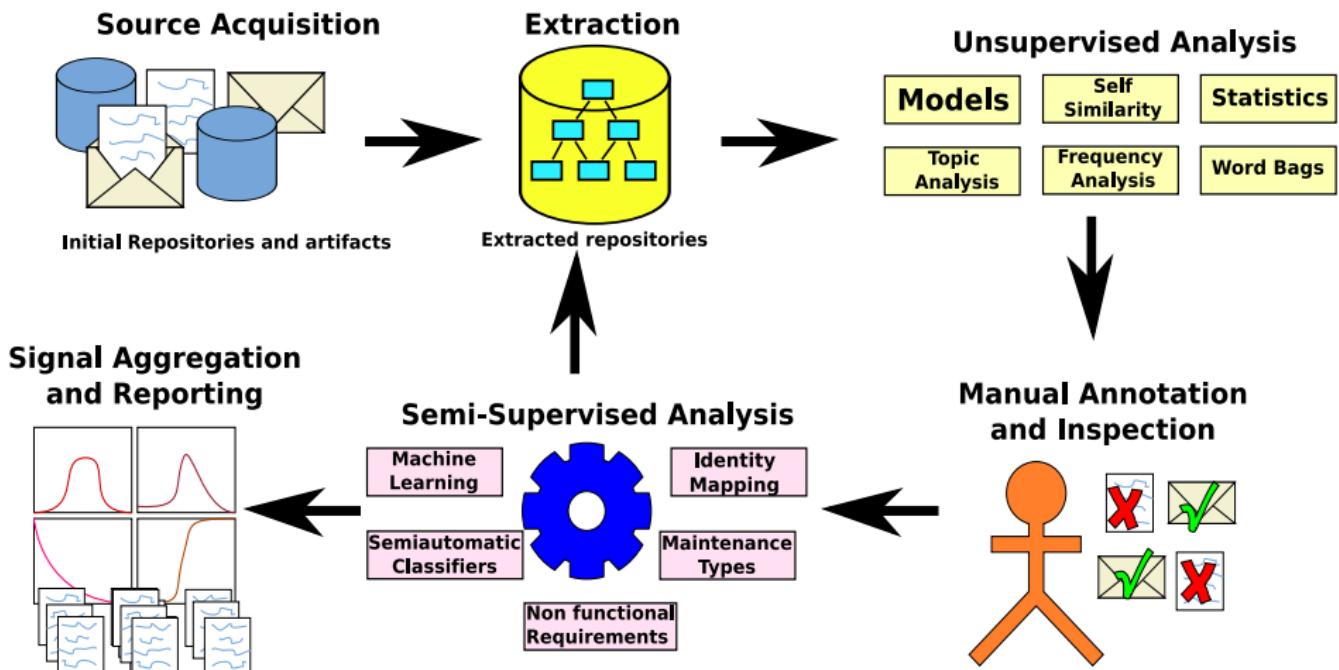
Unsupervised Analysis



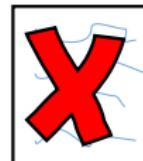
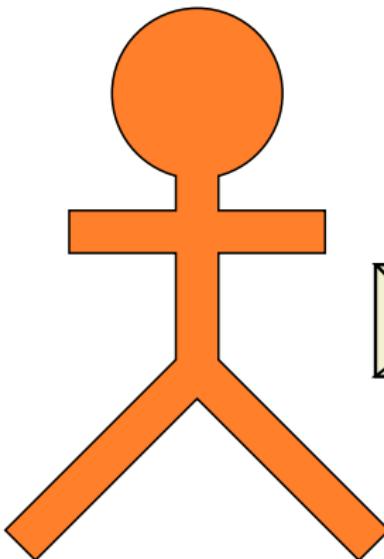
Topic Analysis



Methodology: Recovered Unified Process Views



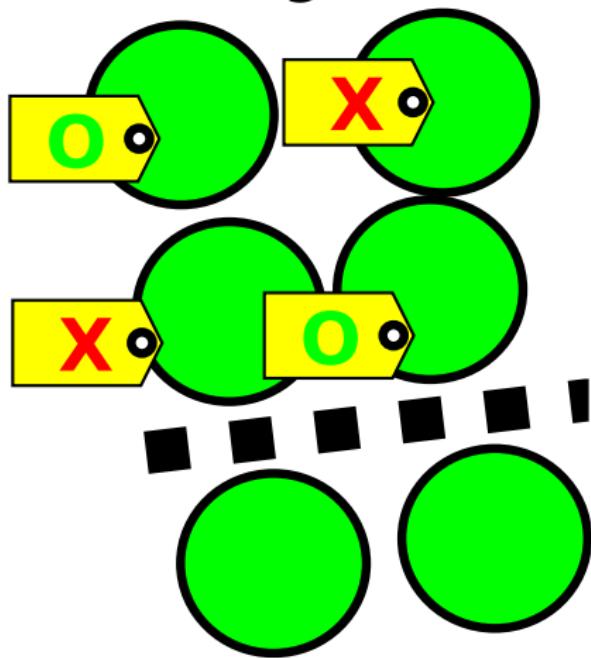
Manual Annotation and Inspection



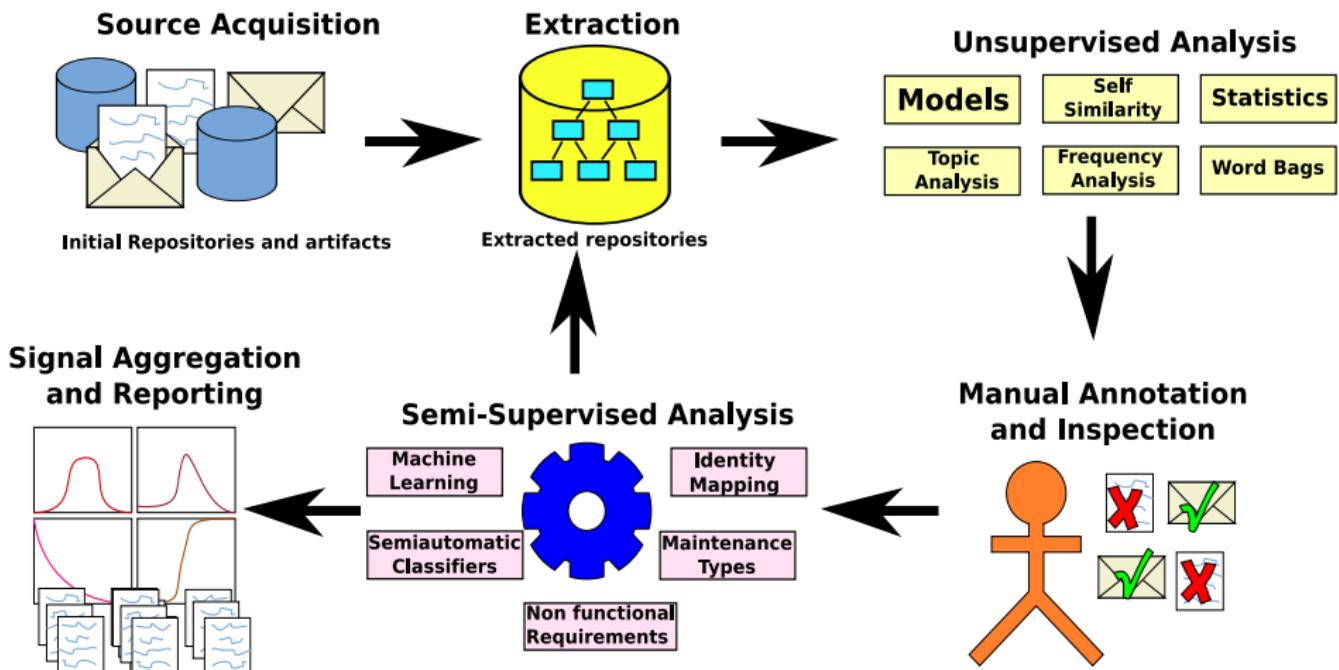
Annotation



Training Set



Methodology: Recovered Unified Process Views

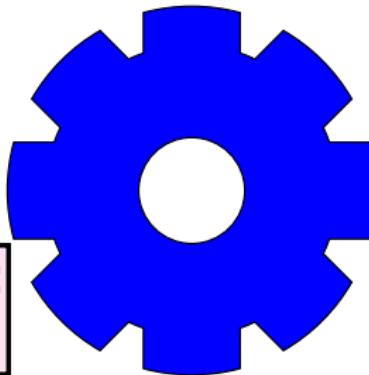




Semi-Supervised Analysis

Machine
Learning

Semiautomatic
Classifiers



Identity
Mapping

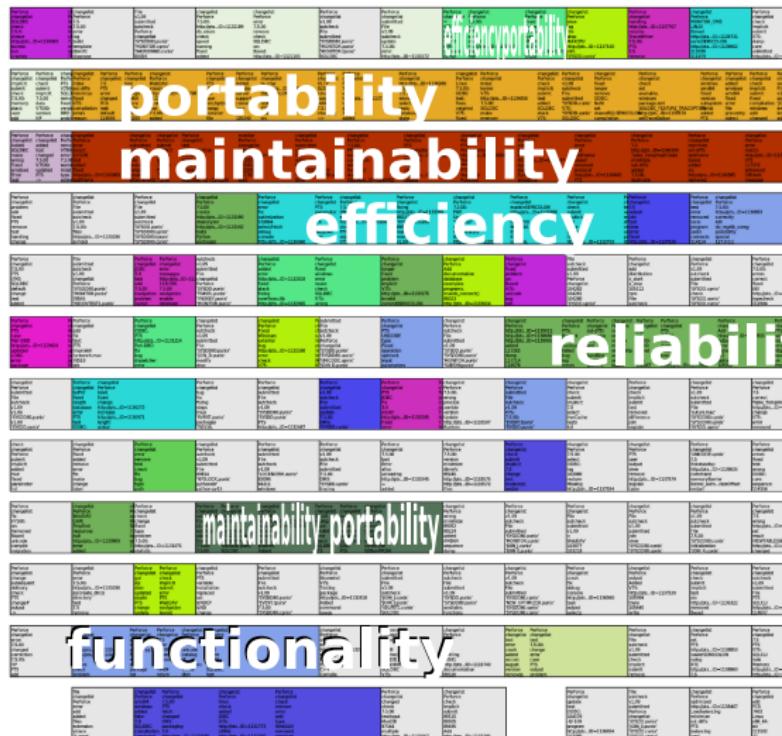
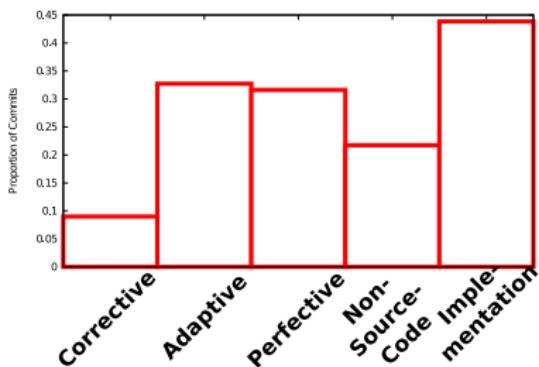
Maintenance
Types

Non functional
Requirements

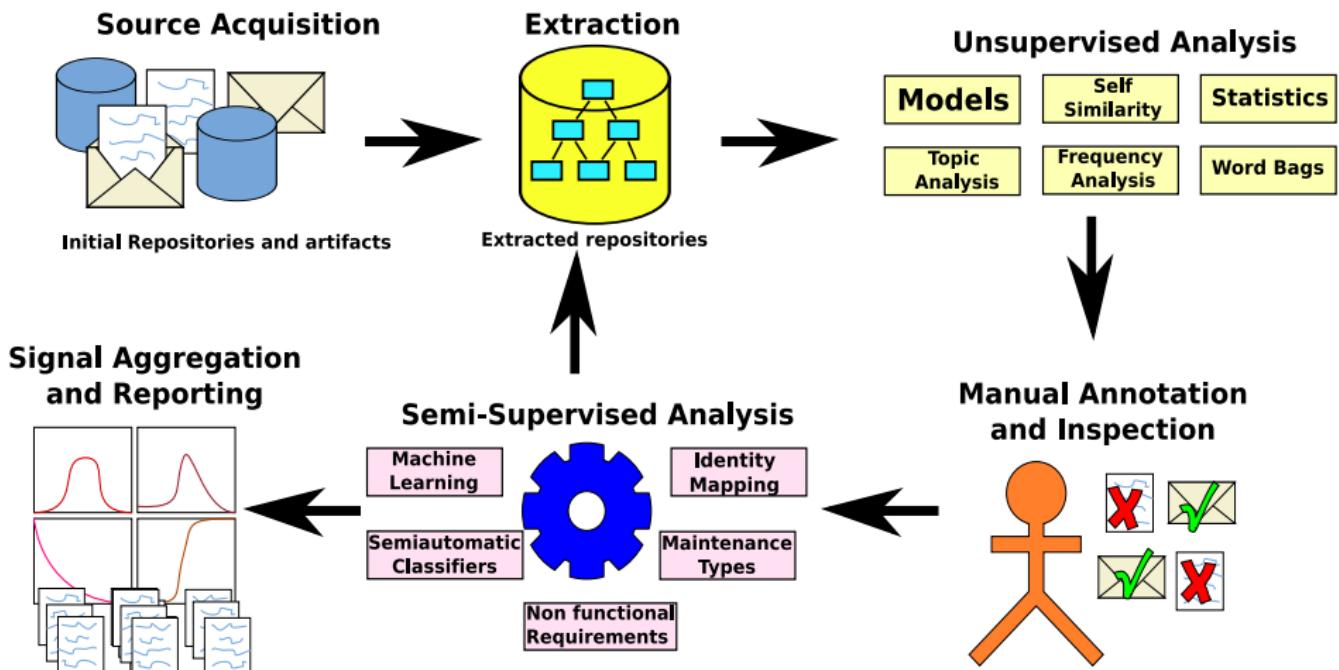
Supervised Analysis

2004 Jun | 2004 Jul | 2004 Aug | 2004 Sep | 2004 Oct | 2004 Nov | 2004 Dec | 2005 Jan | 2005 Jun | 2005 Jul | 2005 Aug | 2005 Oct | 2005 Nov

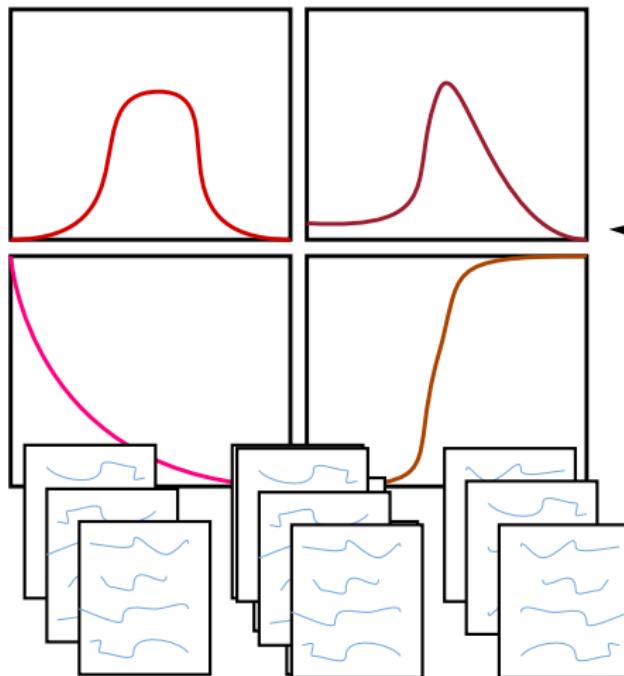
Maintenance Classification



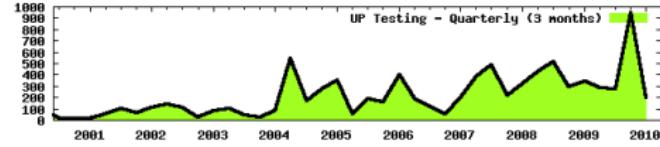
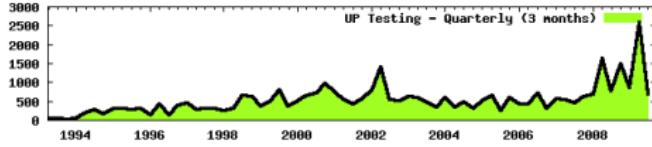
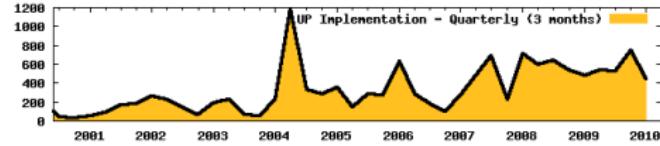
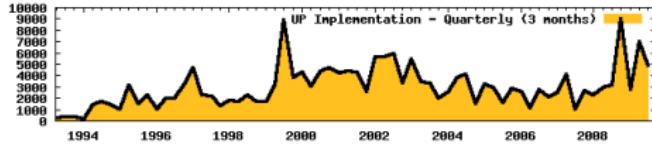
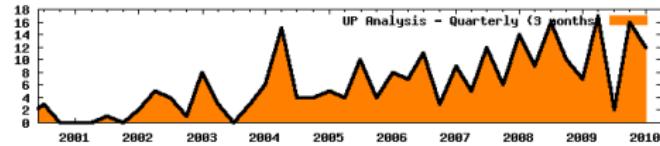
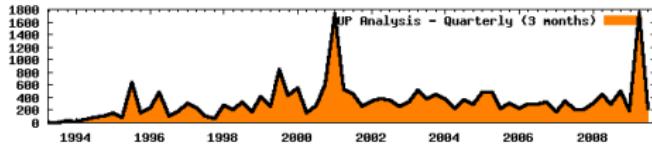
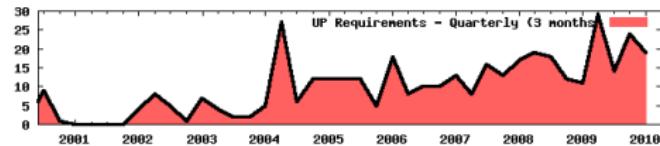
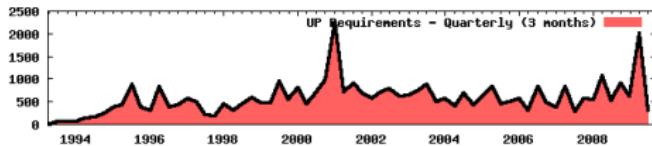
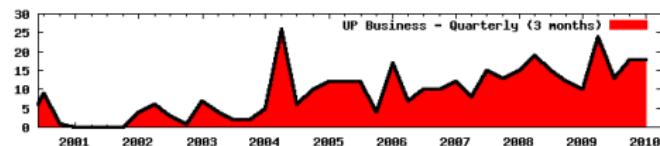
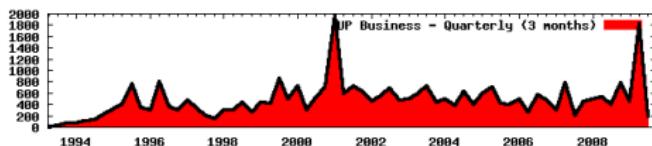
Methodology: Recovered Unified Process Views



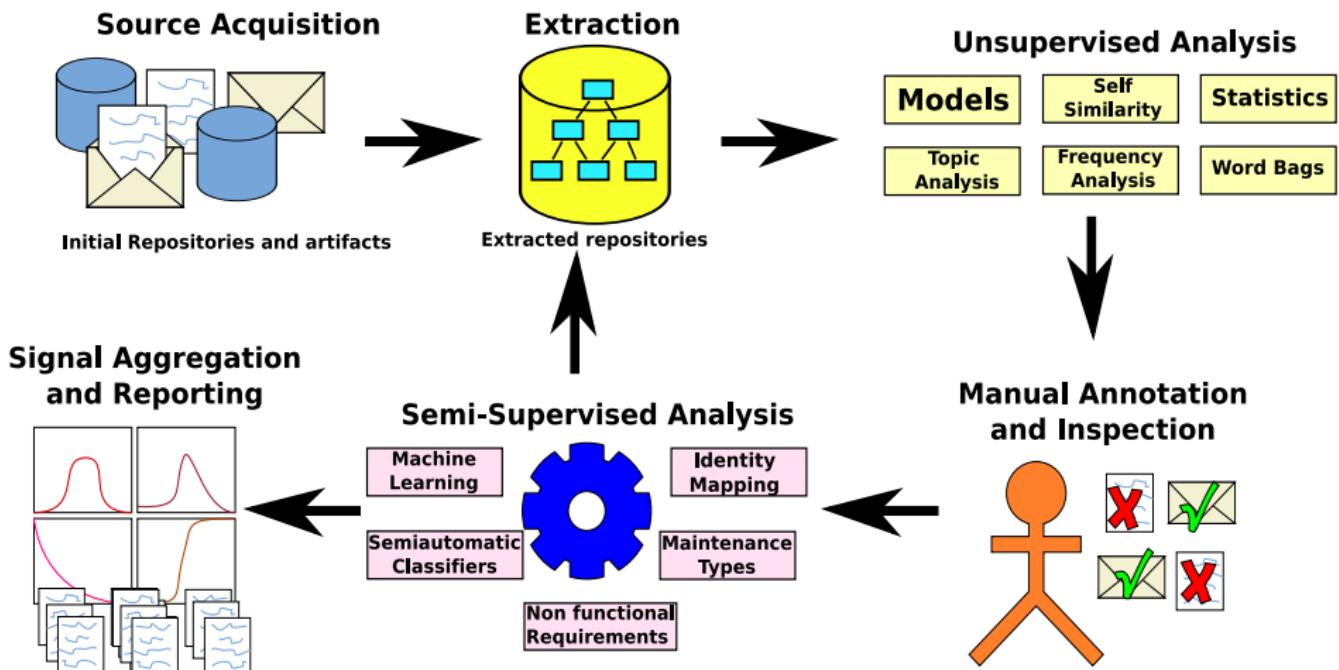
Signal Aggregation and Reporting



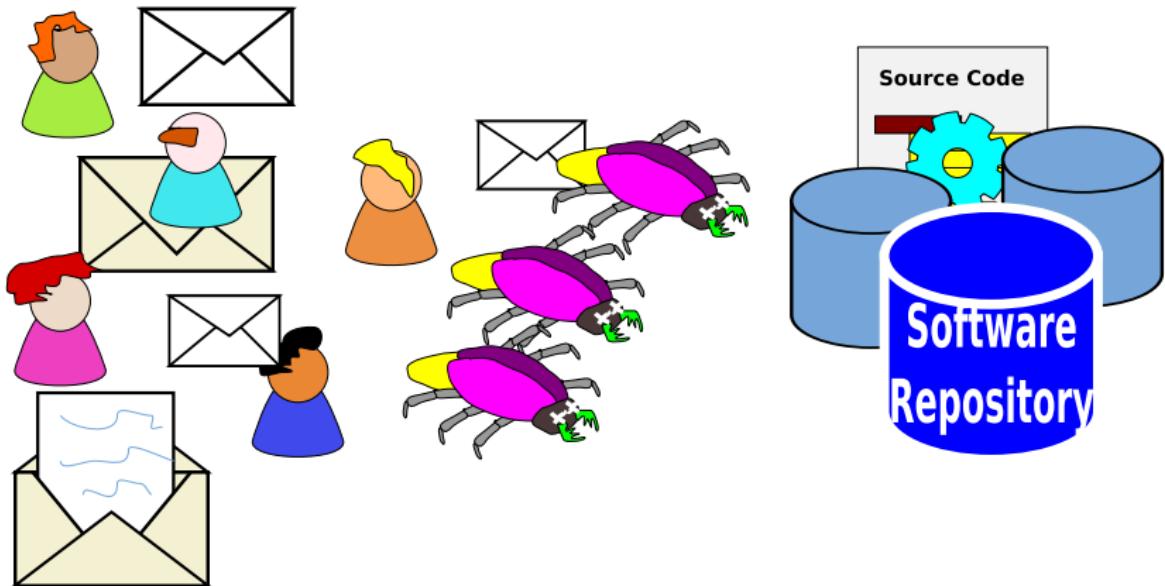
Reporting



Methodology: Recovered Unified Process Views



Source Acquisition



Initial Repositories and artifacts

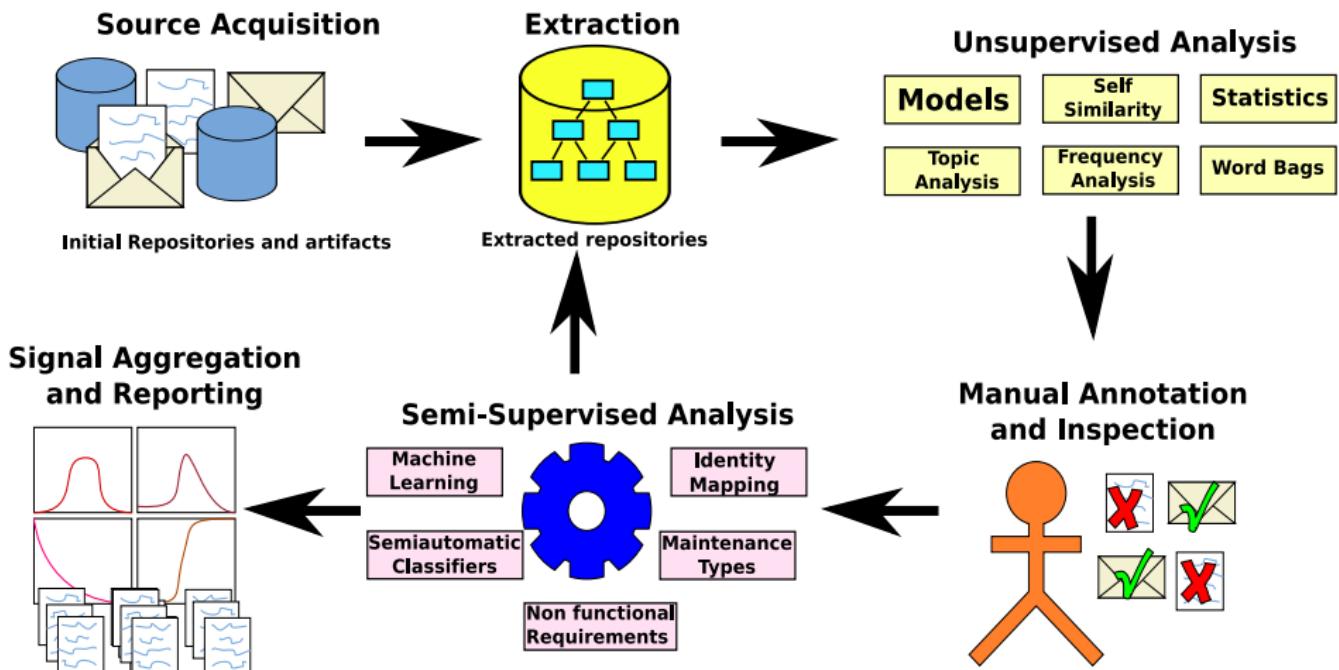
Source Acquisition

- Documentation
 - Wikis
 - Websites
 - Docs in source control
 - Design
 - Requirements

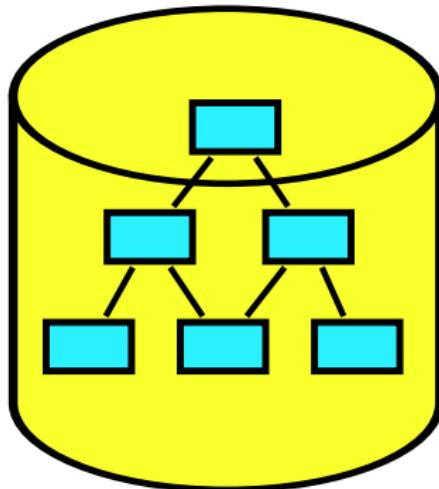
Source Acquisition

- Discussions
 - Mailing lists
 - Forums
 - Instant Messaging
 - IRC

Methodology: Recovered Unified Process Views

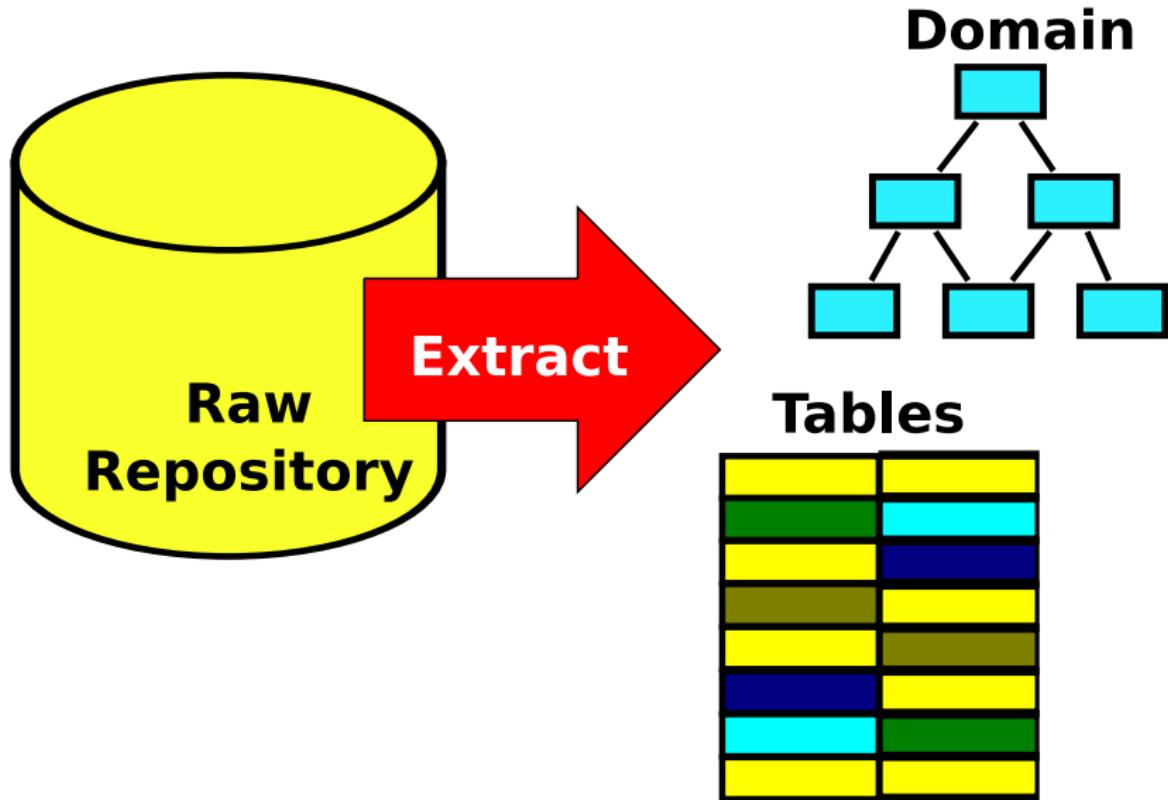


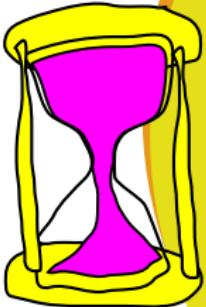
Extraction



Extracted repositories

Extraction



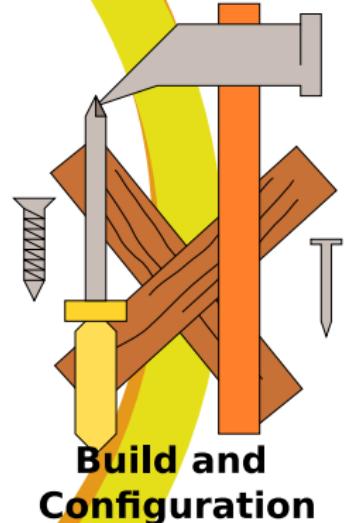


time

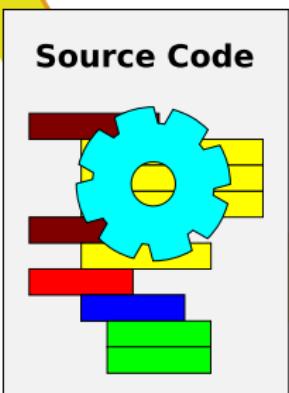


Revisions

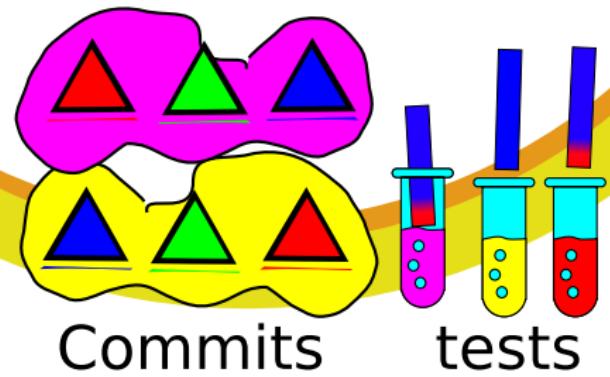
authors



Build and Configuration



Source Code



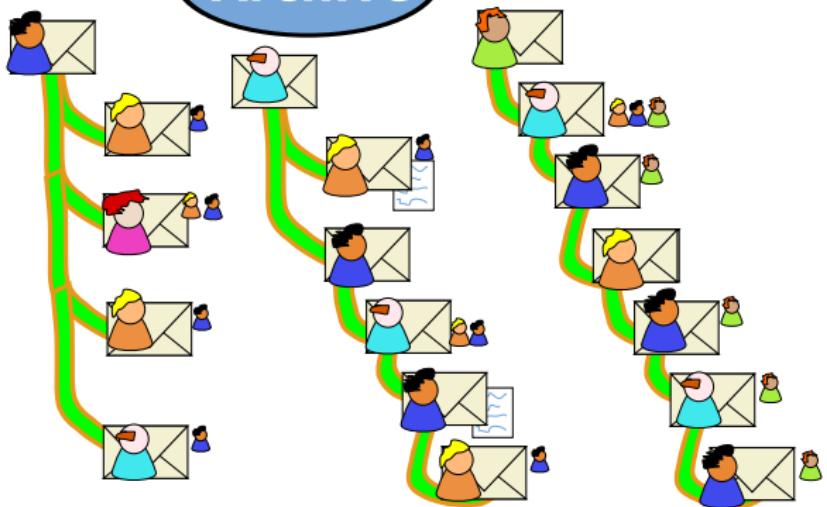
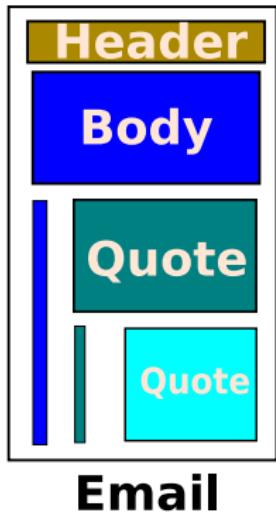
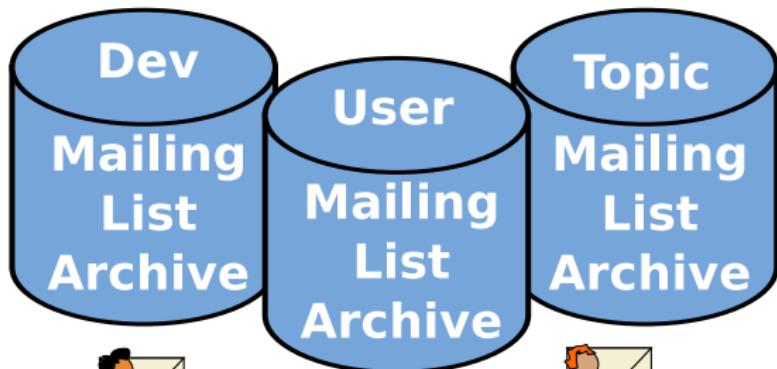
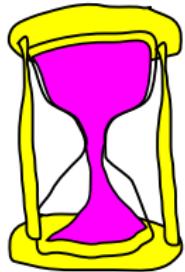
Commits

tests

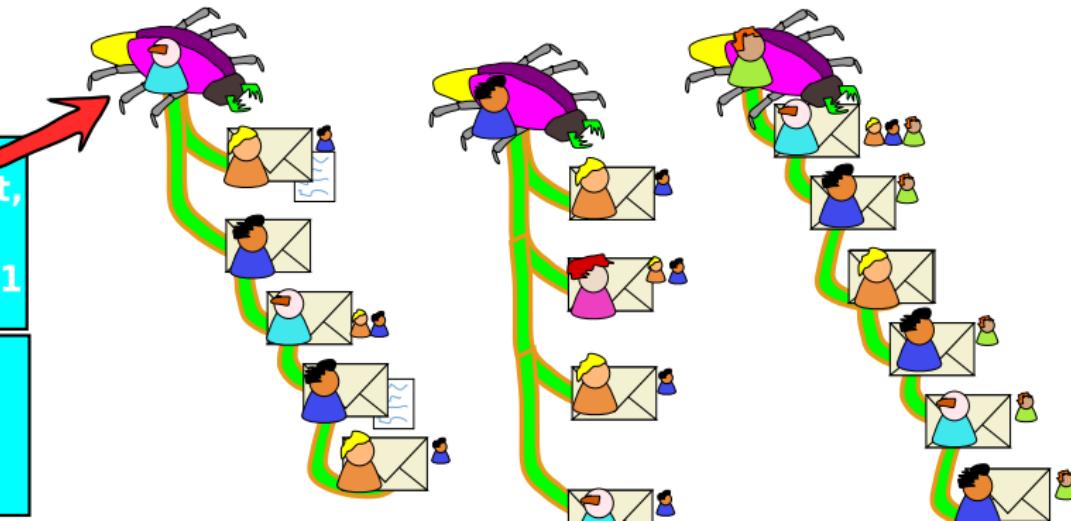
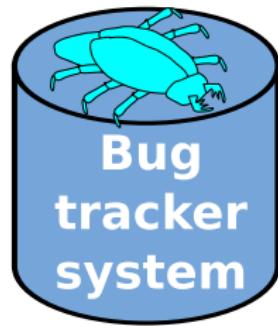
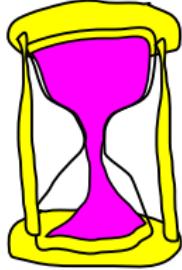


Documentation

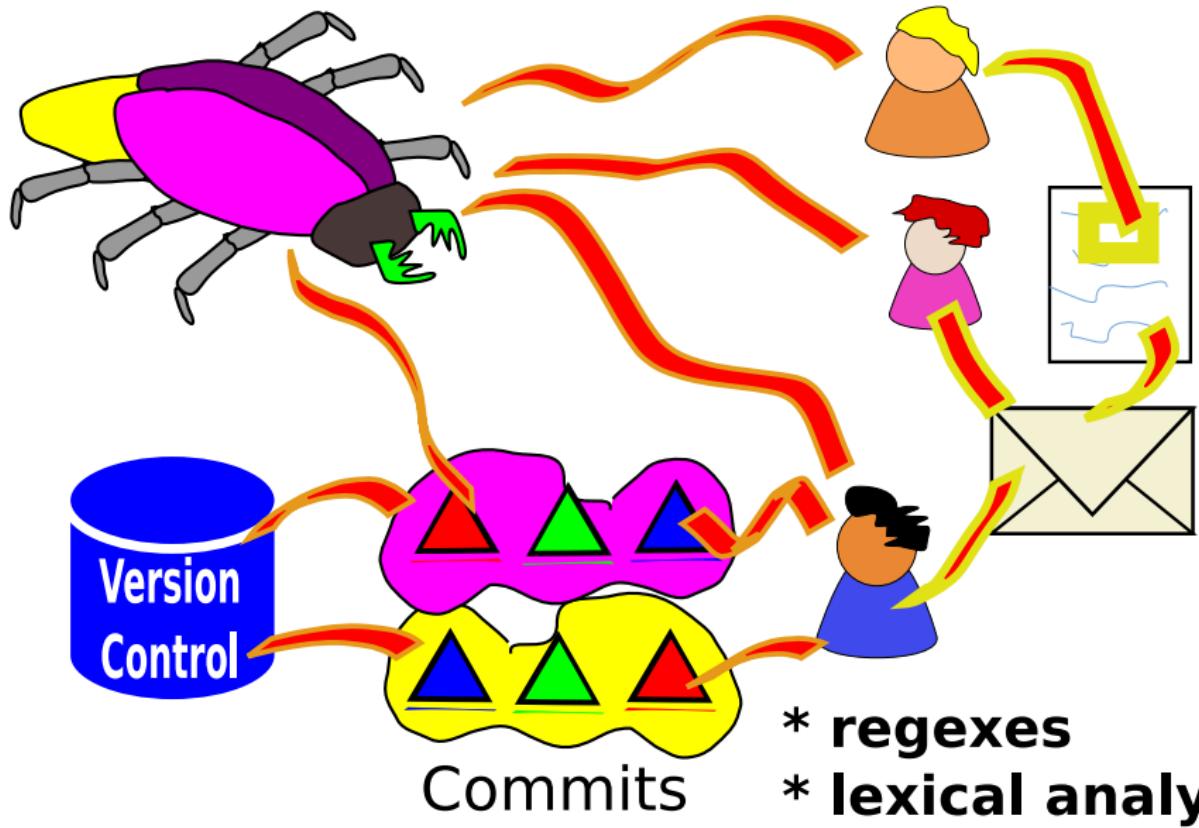
Extraction: Mailing list archives



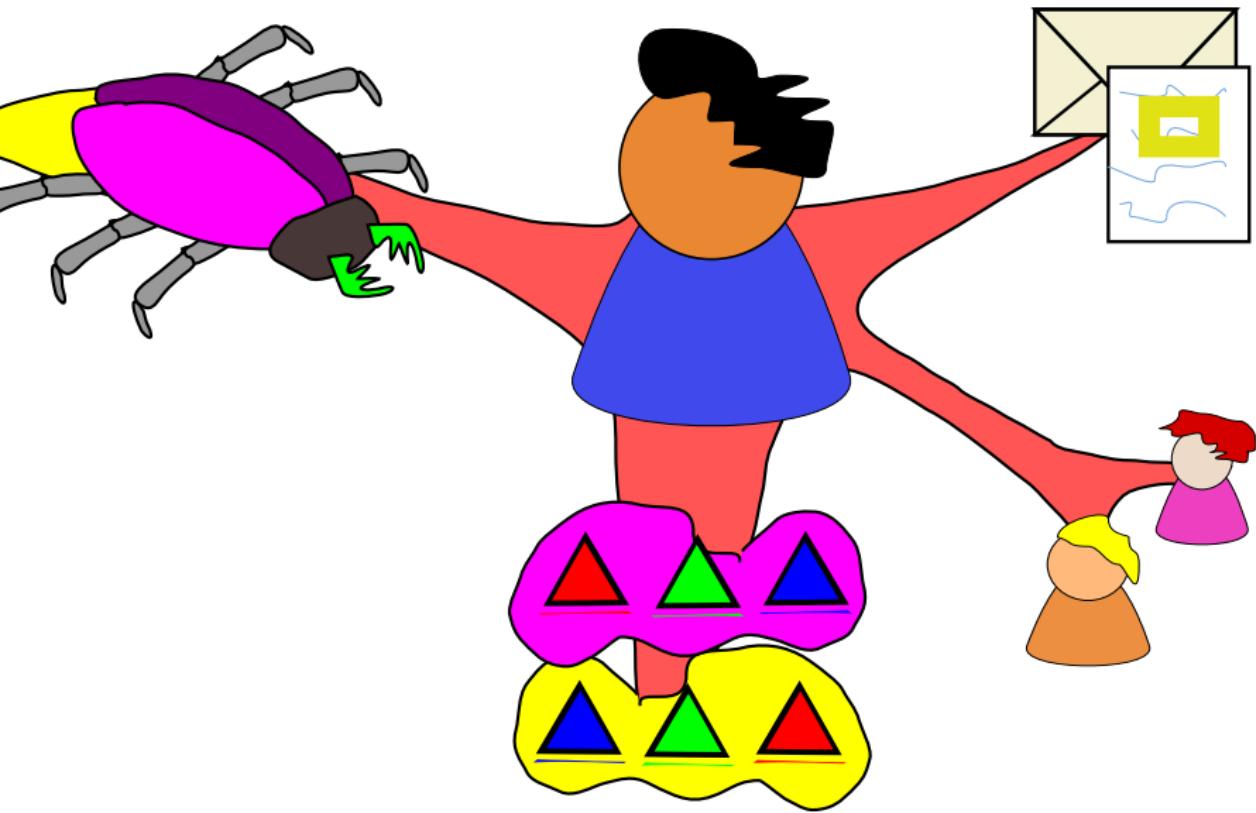
Extraction: Bug trackers



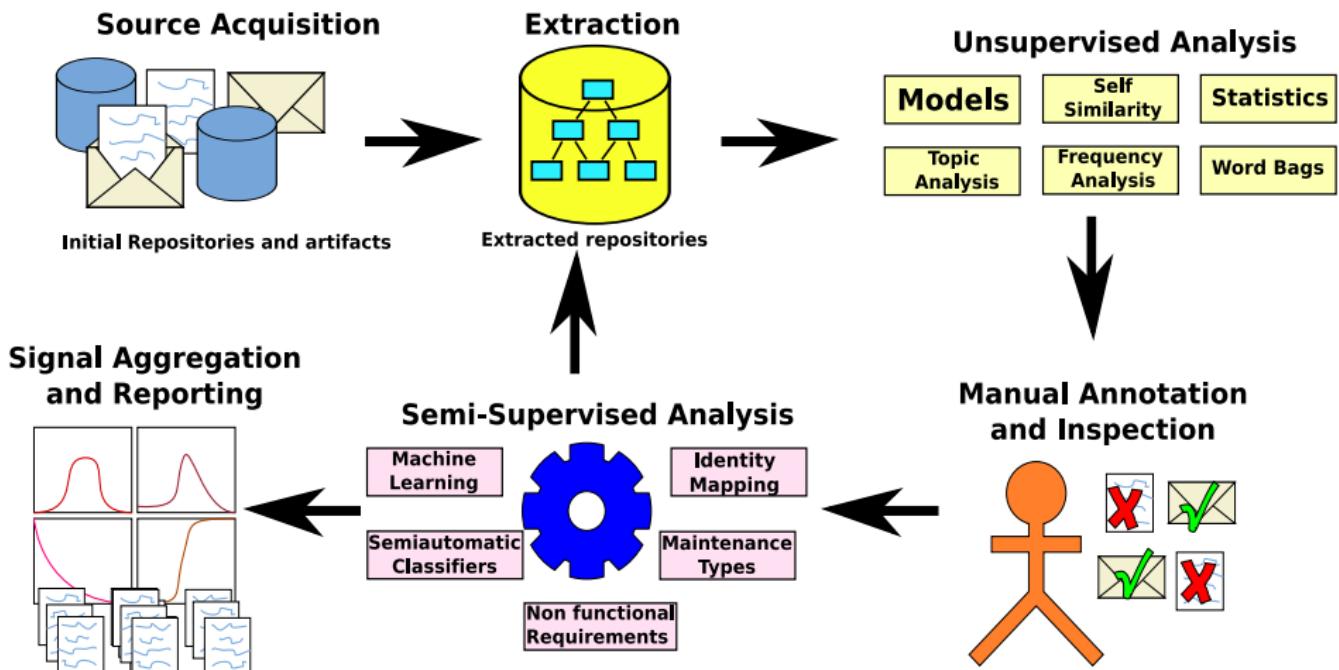
Extraction: Traceability



Extraction: People



Methodology: Recovered Unified Process Views



Unsupervised Analysis

Models

**Self
Similarity**

Statistics

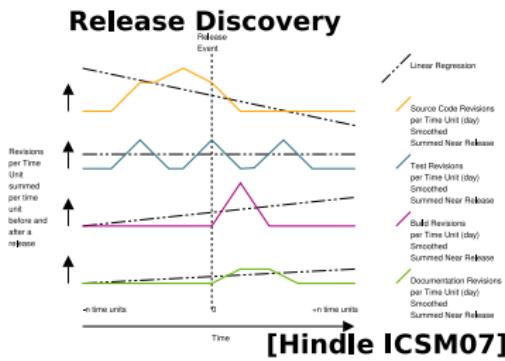
**Topic
Analysis**

**Frequency
Analysis**

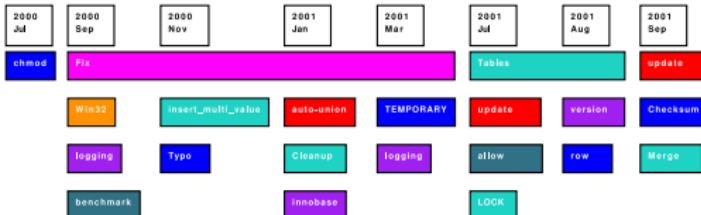
Word Bags



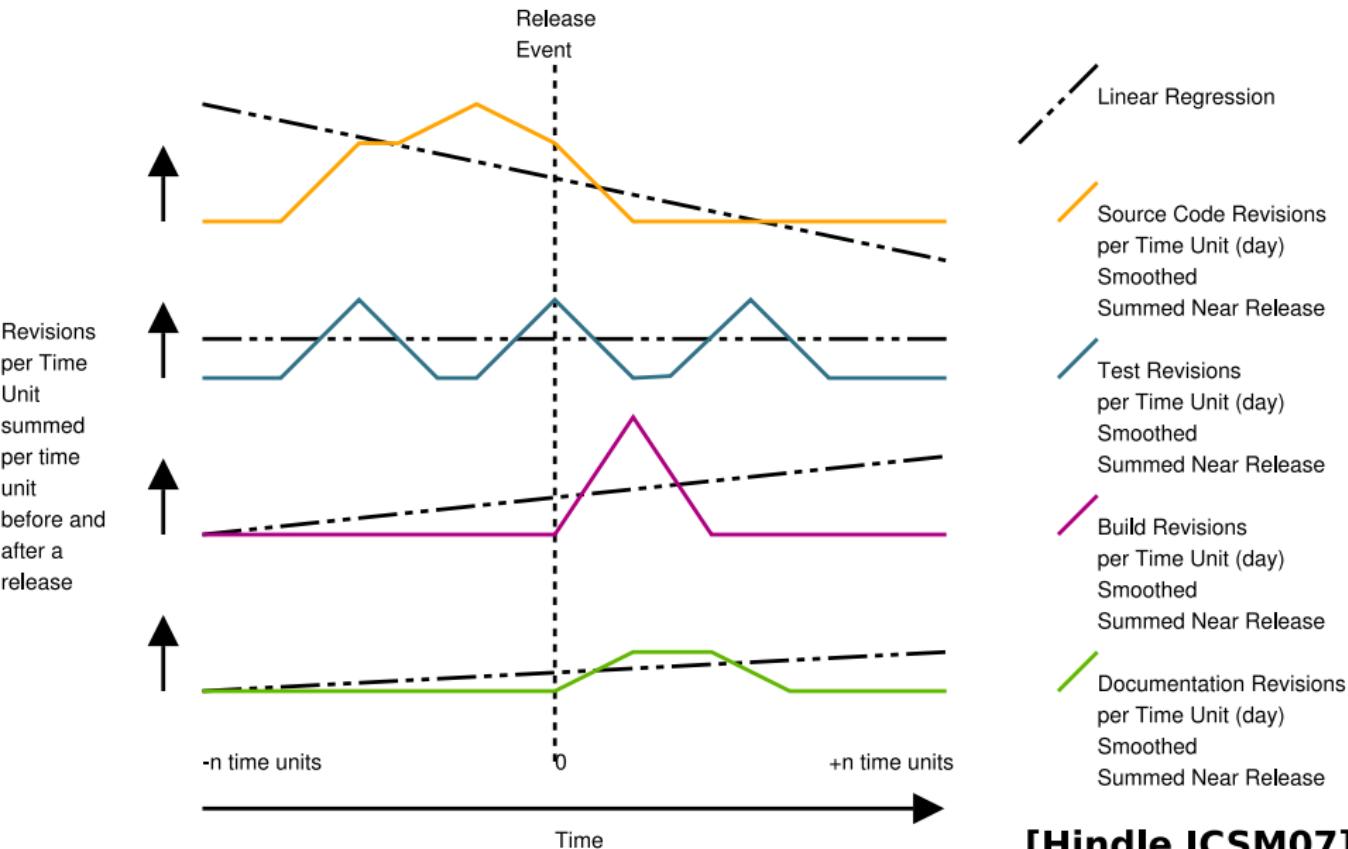
Unsupervised Analysis



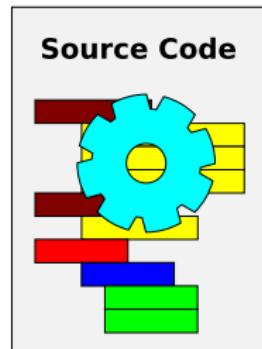
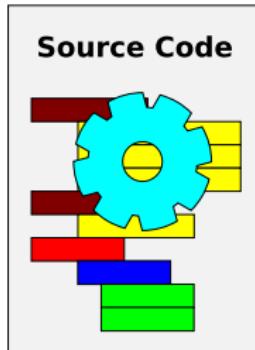
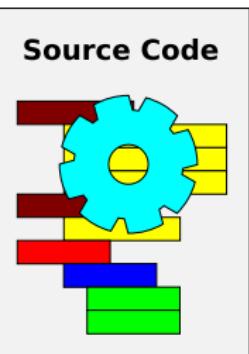
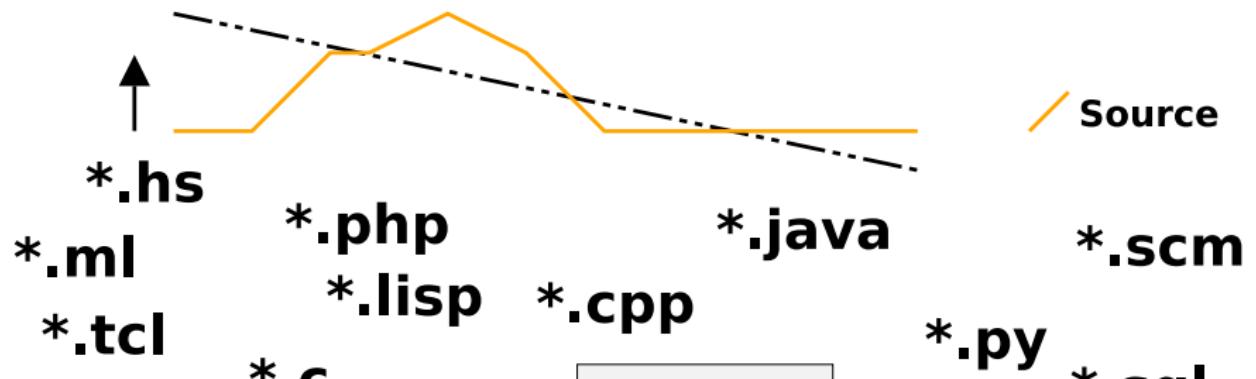
Topic Analysis



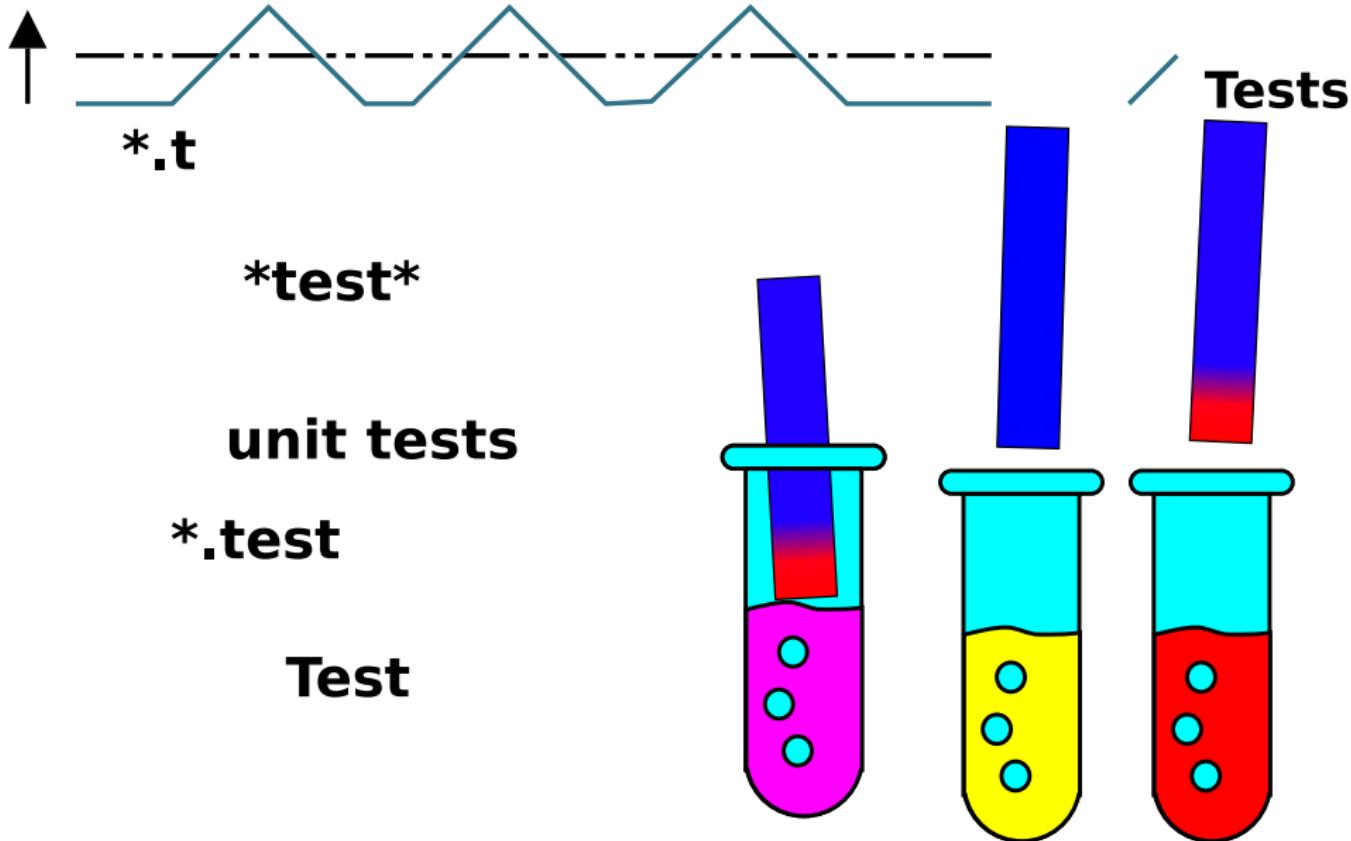
Unsupervised Analysis: STBD



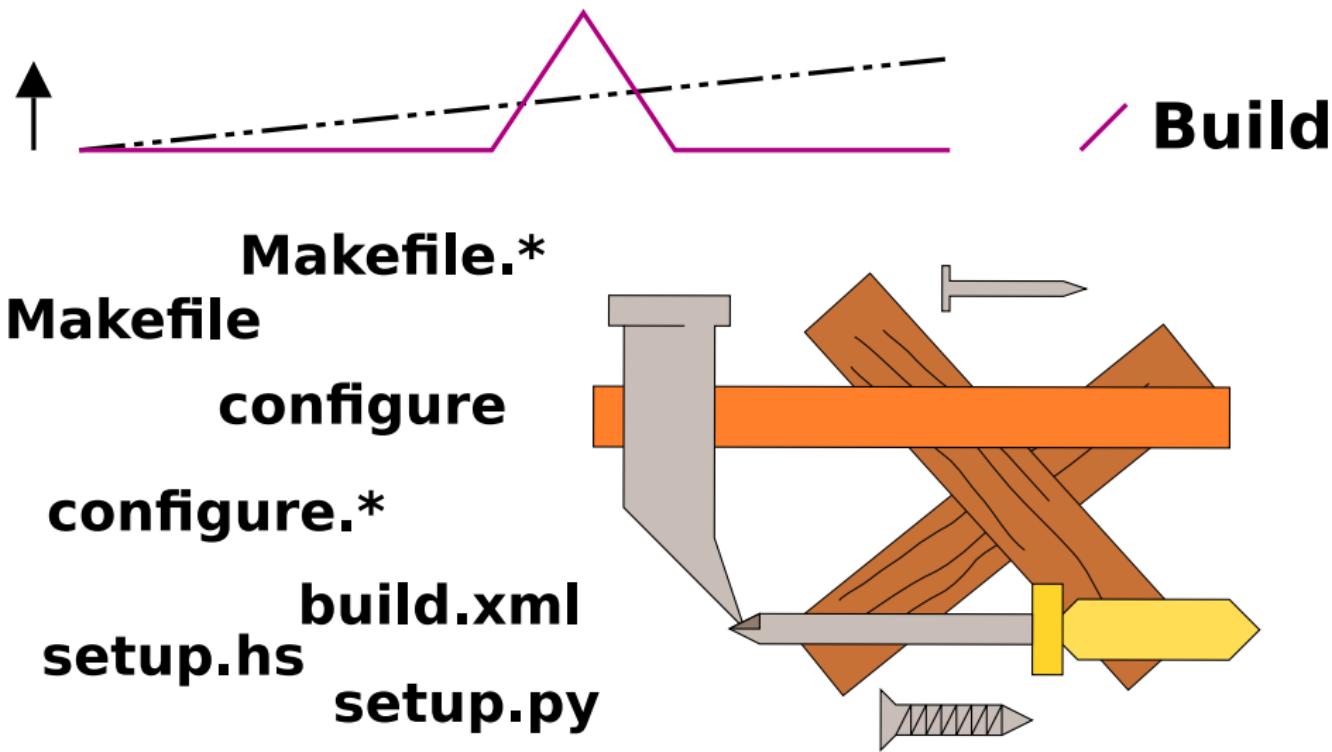
Unsupervised Analysis: Source



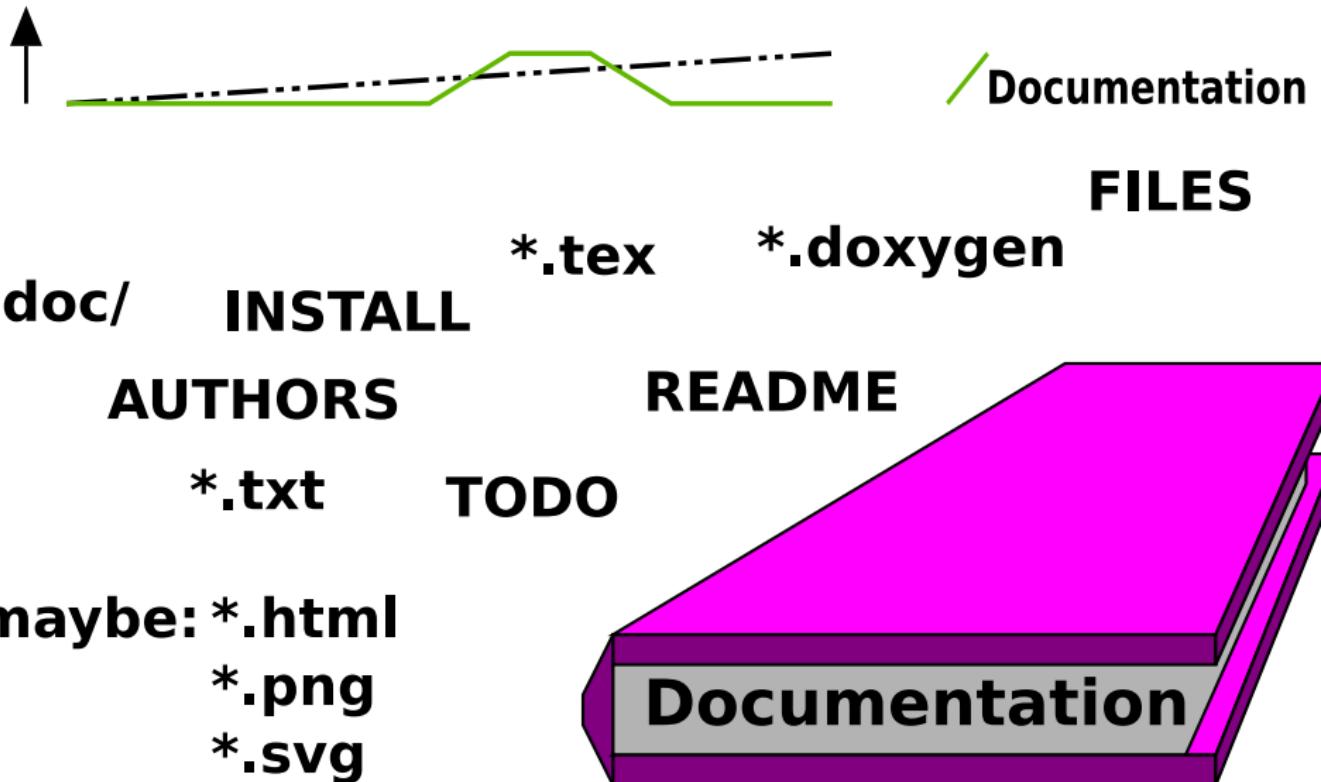
Unsupervised Analysis: Testing



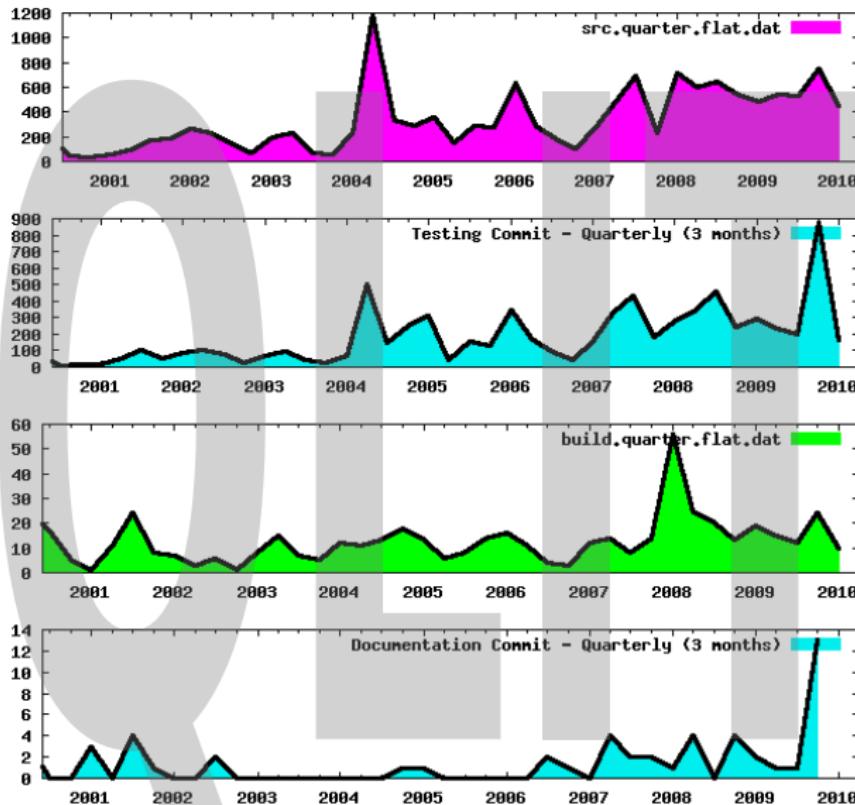
Unsupervised Analysis: Build files

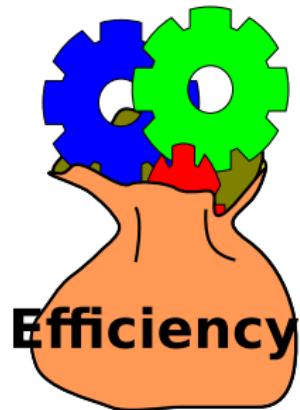
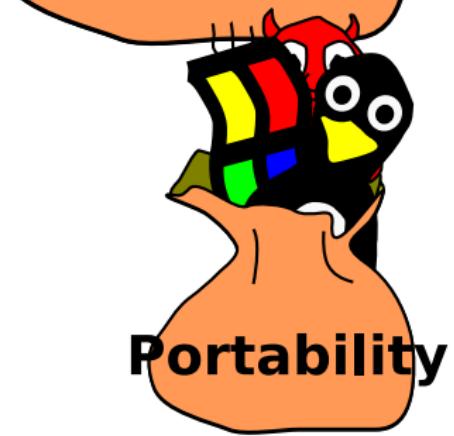


Unsupervised Analysis: Documentation



Unsupervised Analysis: STBD applied





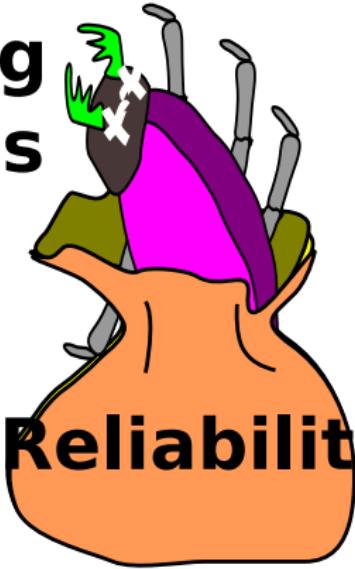


Word Bag Examples

Portability

portability
transferability
interoperability
documentation
internationalization
i18n

...

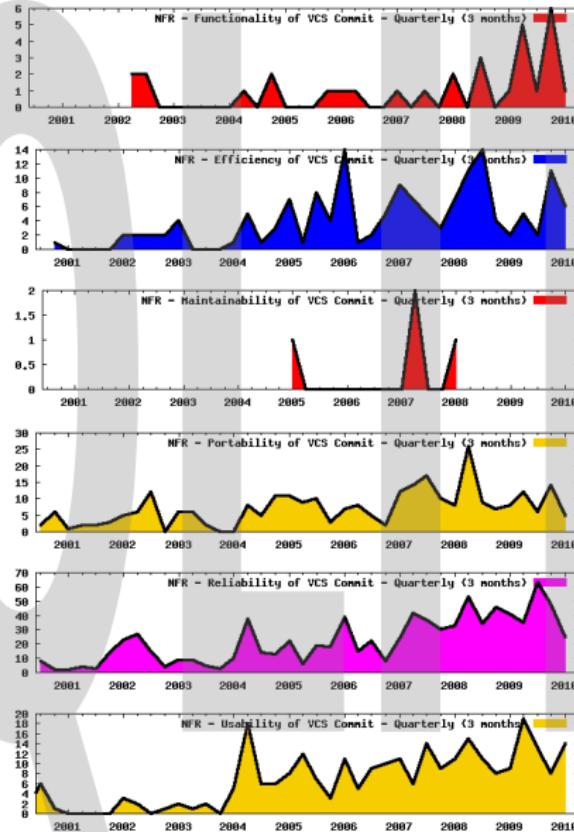


Reliability

reliability
failure
error
redundancy
fails
bug

...

Unsupervised Analysis: Word Bag Applied



Please excuse the tangent



Developer Topics

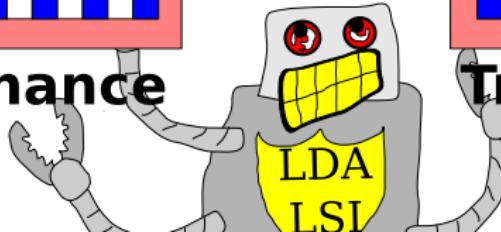


Maintenance

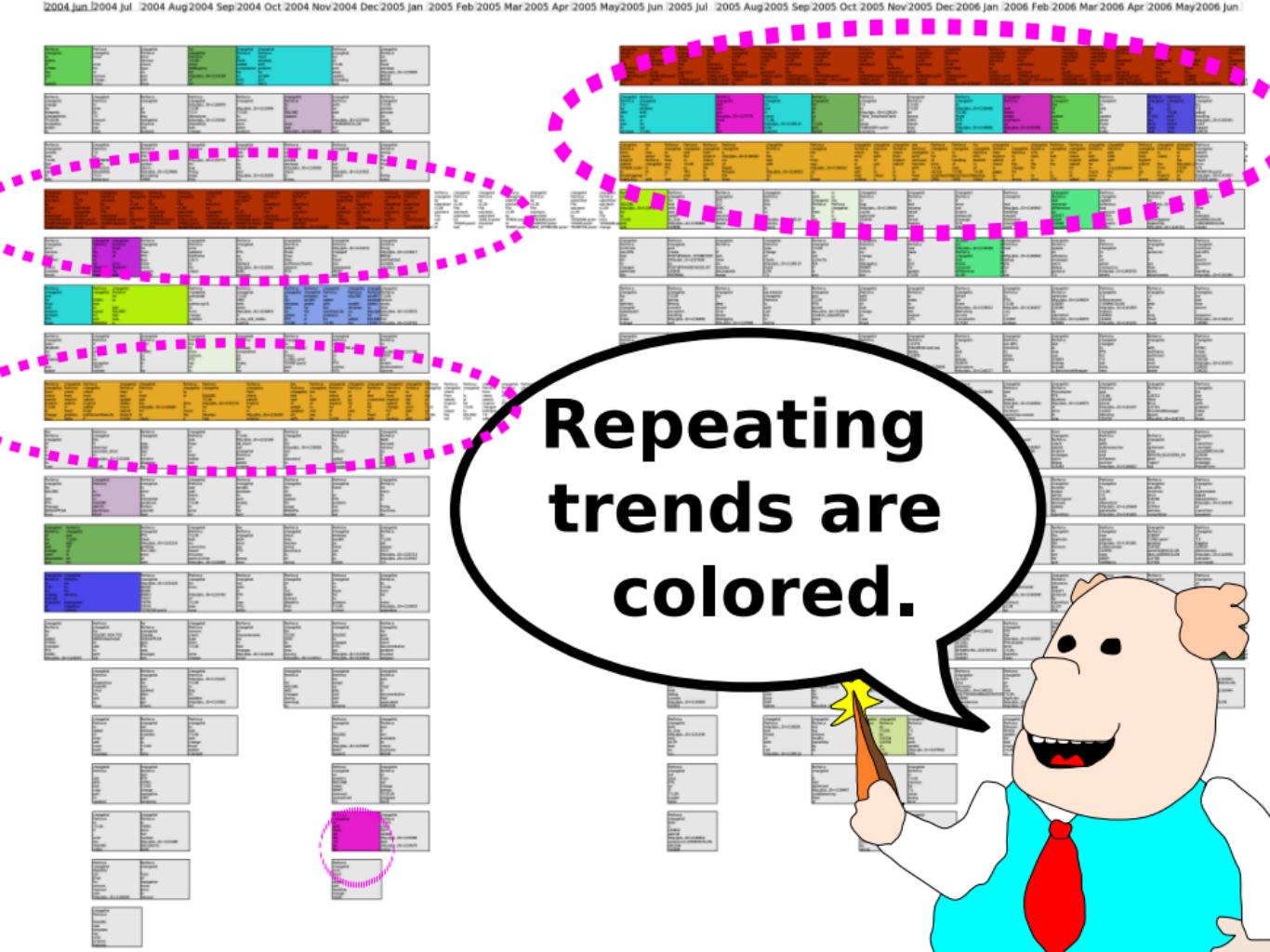


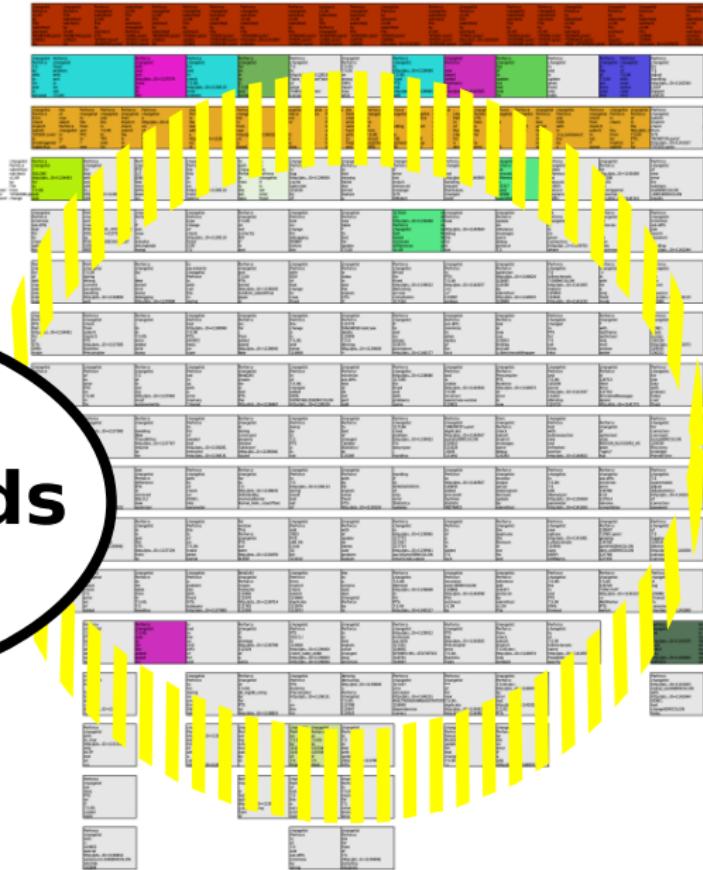
Transactions

purpose?



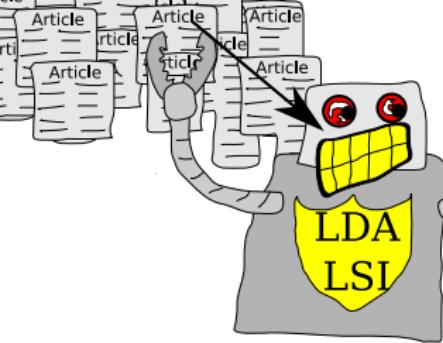






Local
unique trends
are grey

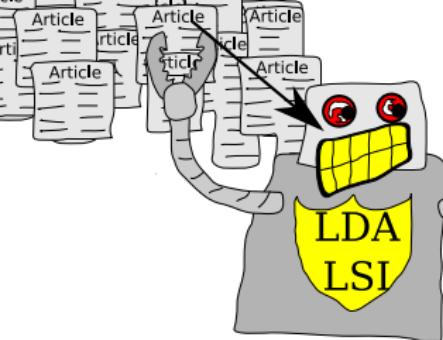




Here are two topics. I
don't know what they are
about!

Topic 1 Topic 2

* win32	* version
* backport	* bump
* fix	* number
* bug	* up
* ...	* ...



Here are two topics. I
don't know what they are
about!

Topic 1

- * win32
- * backport
- * fix
- * bug
- * ...

Topic 2

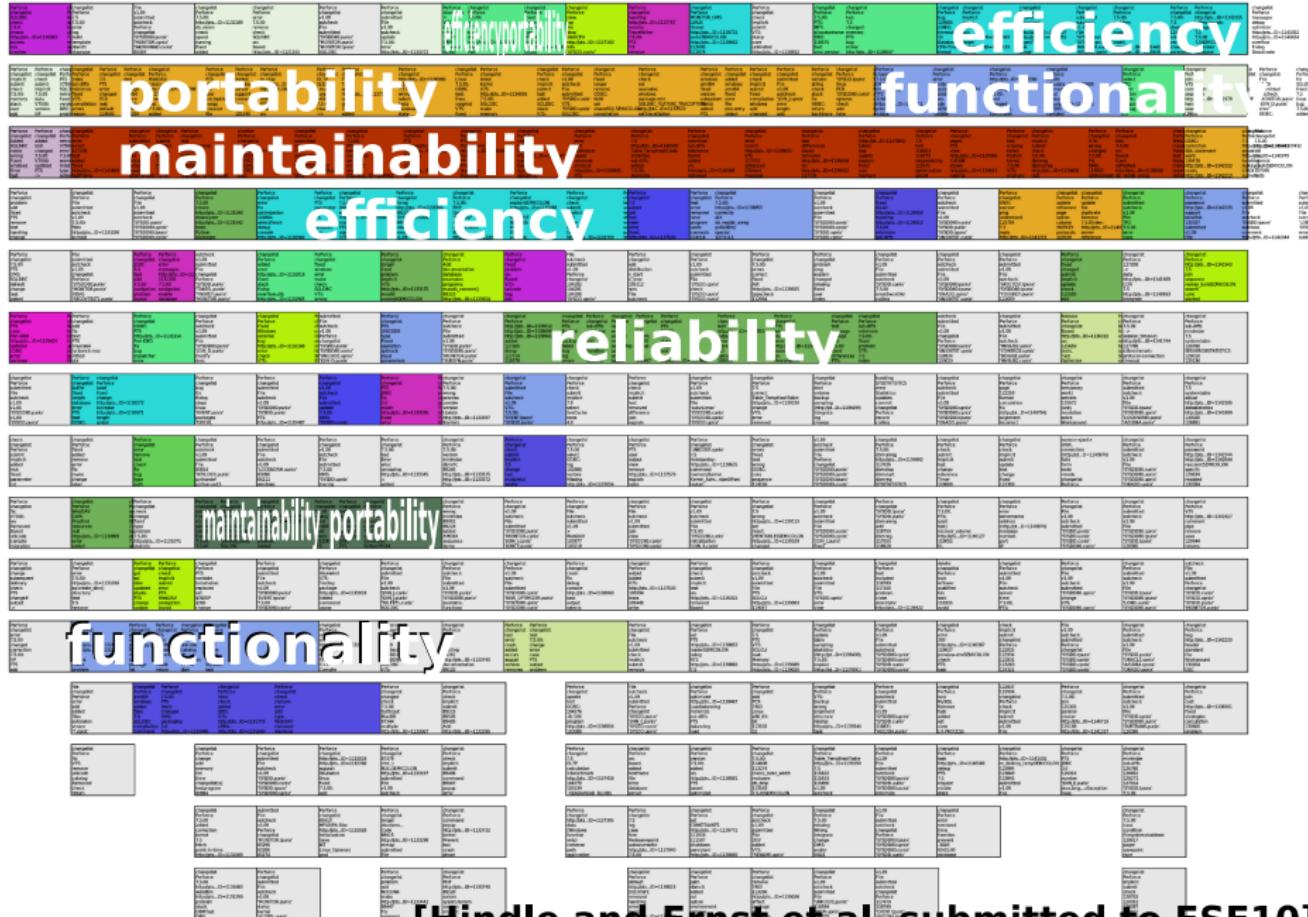
- * version
- * bump
- * number
- * up
- * ...

These word lists
might be: **Windows**
backport and
versioning!

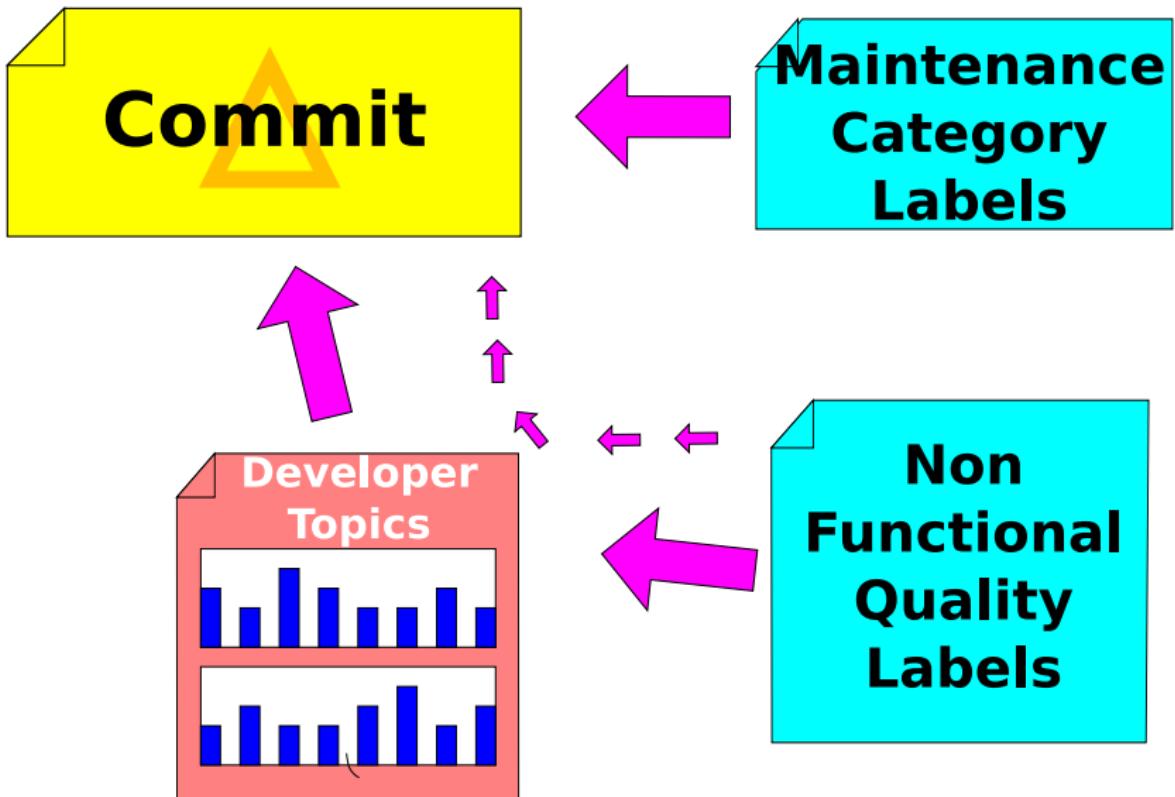


Label topics by quality

- Non functional requirements
 - Maintainability
 - Functionality
 - Portability
 - Efficiency
 - Usability
 - Reliability



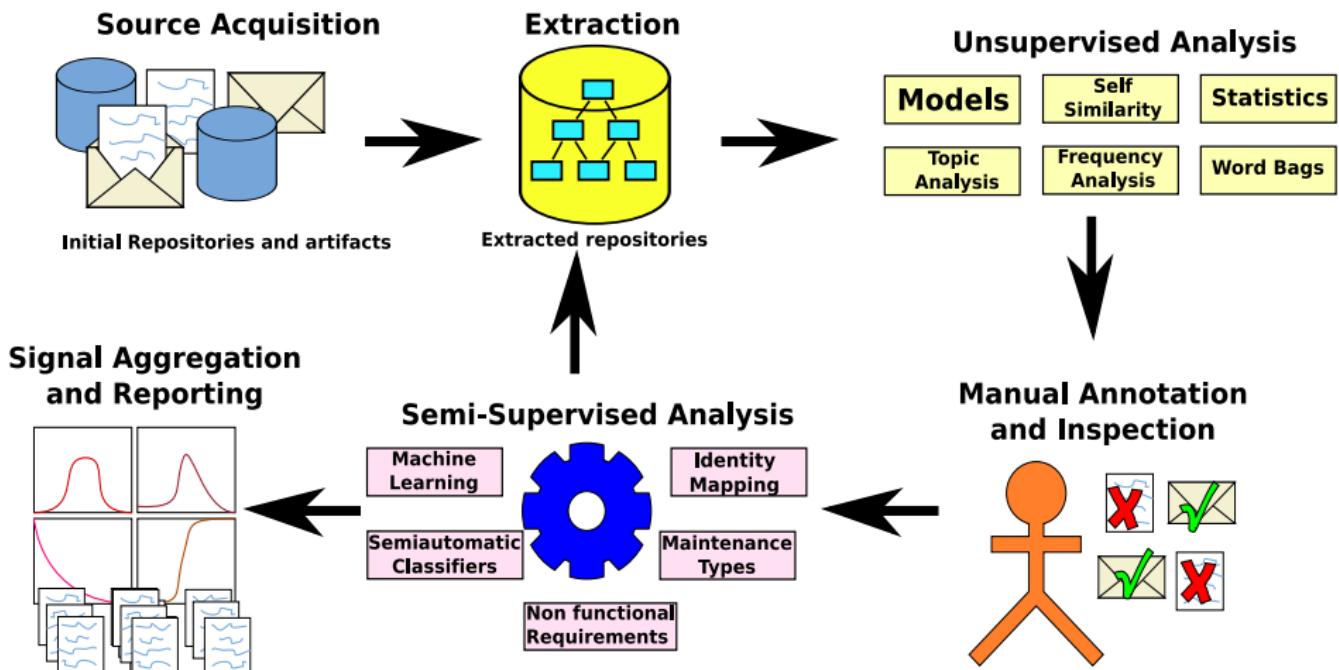
Automatic Labelling



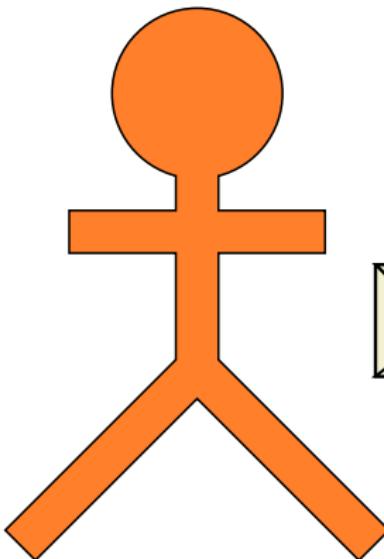
That wasn't too bad...



Methodology: Recovered Unified Process Views



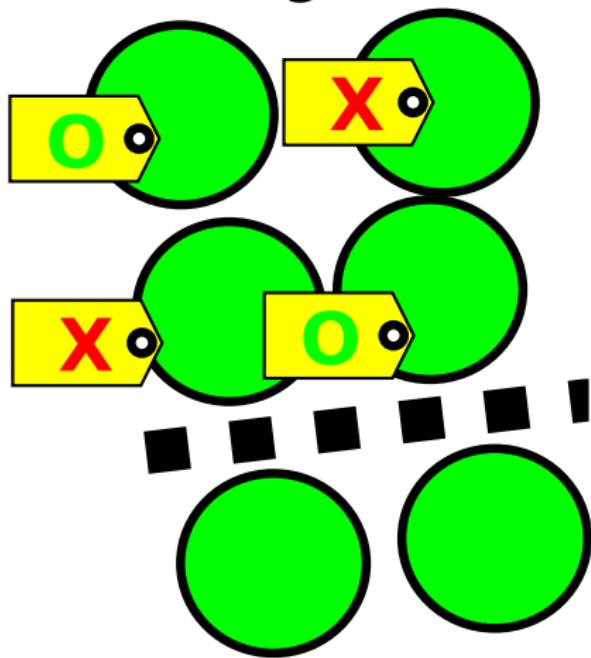
Manual Annotation and Inspection



Annotation



Training Set



Annotation: Stop Words

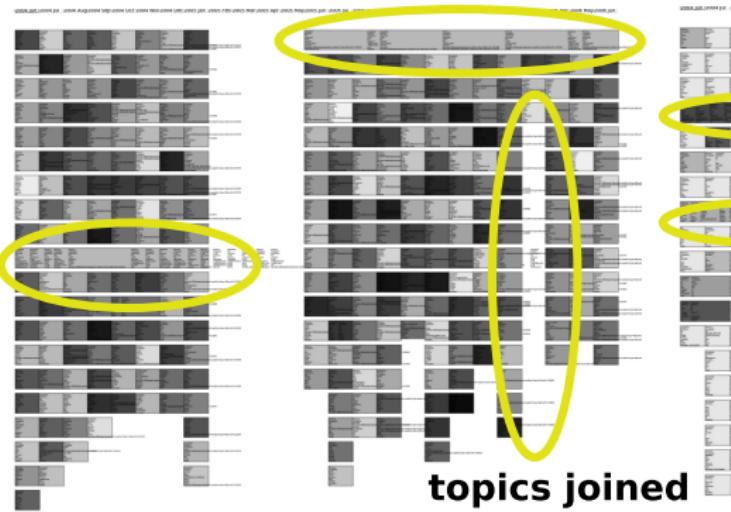


**Used in topic analysis
or to reduce # of
features for learners.**

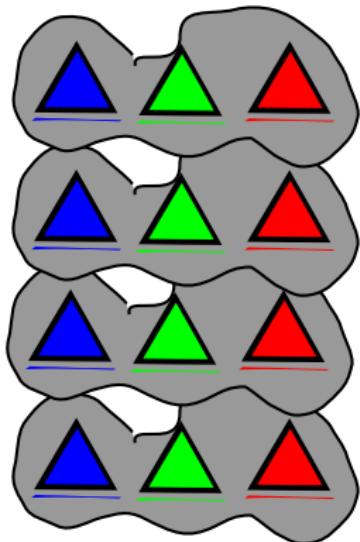
perhaps
clearly
between
done them
who because
haven't move
in asking exam
nevertheless doe
sensible our some
elsewhere upon ask
beforehand ie found
anywhere it contain
everywhere detail
need association
specifying
con d: for

Annotation: Stop Words

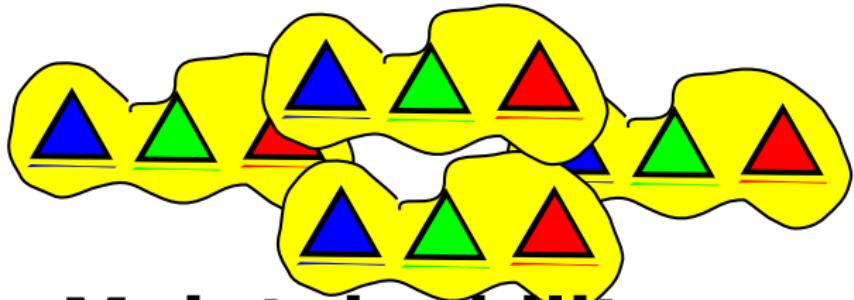
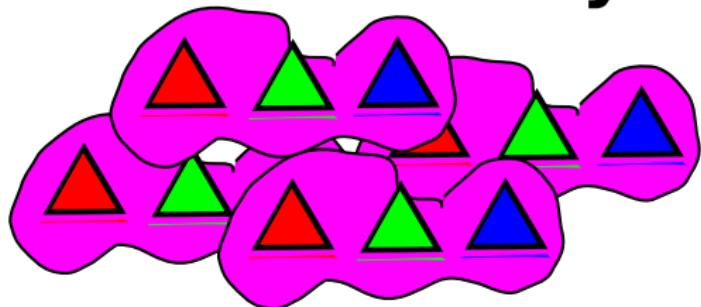
MaxDB 7.500 Case Study



Annotation: Training Sets

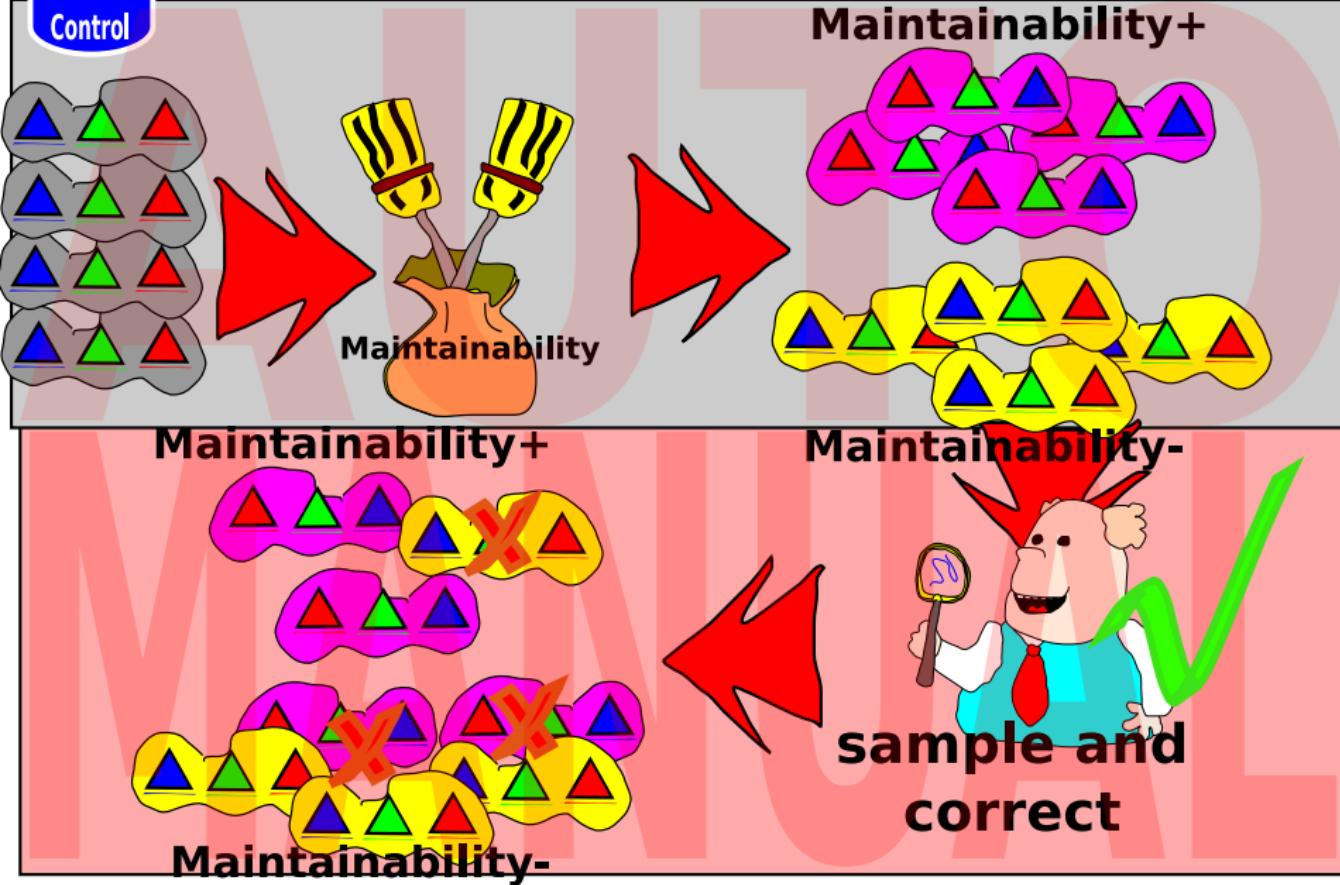


Maintainability+

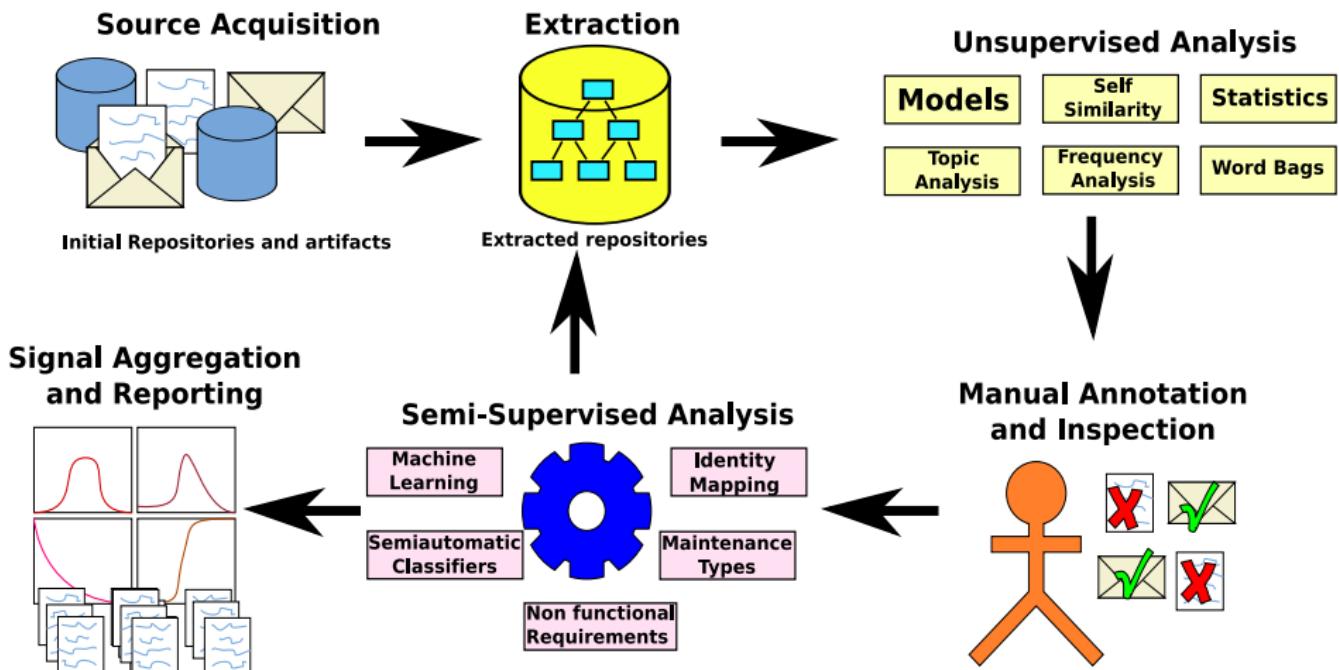


Maintainability-

Annotation: Training Sets



Methodology: Recovered Unified Process Views

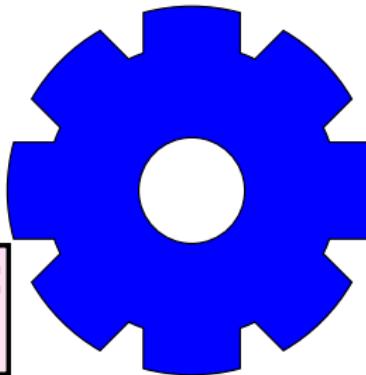




Semi-Supervised Analysis

Machine
Learning

Semiautomatic
Classifiers



Identity
Mapping

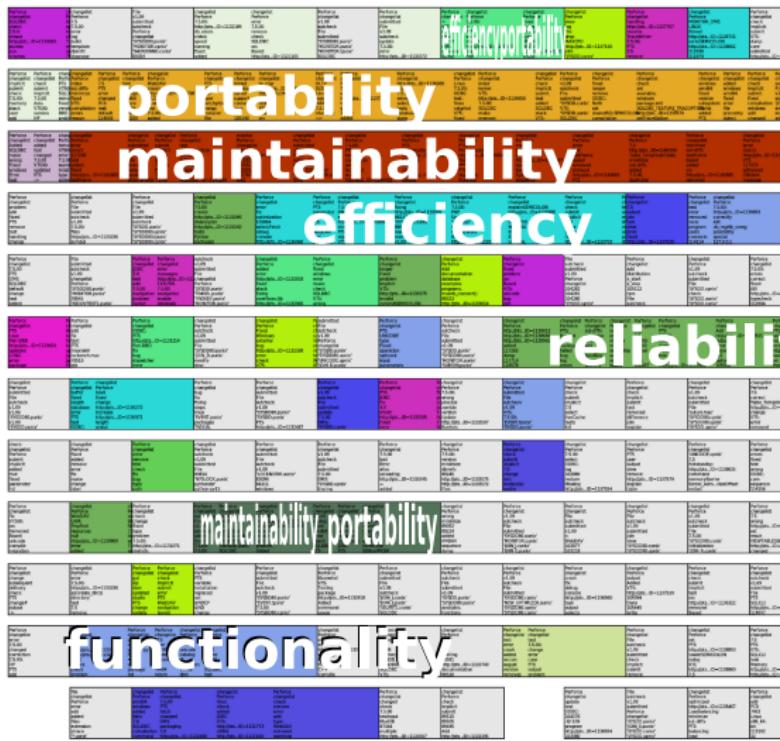
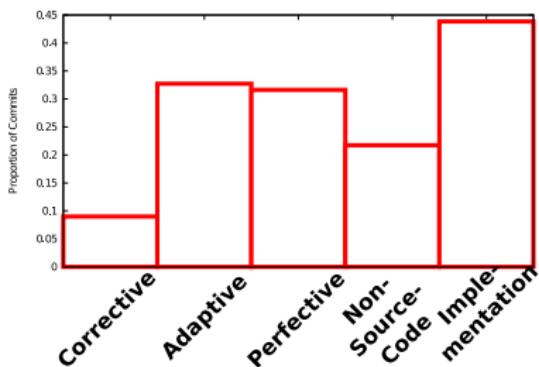
Maintenance
Types

Non functional
Requirements

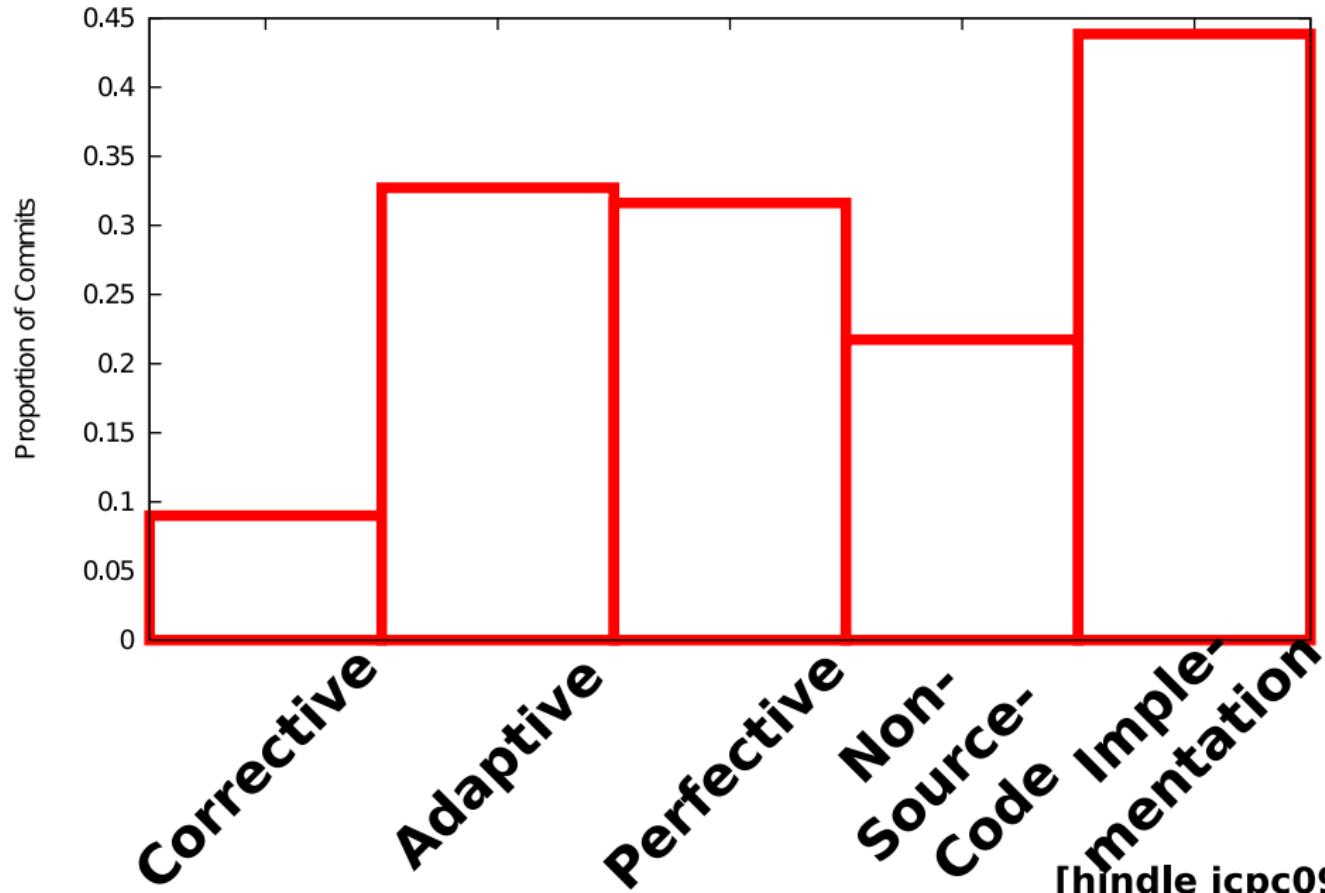
Supervised Analysis

2004 Jun | 2004 Jul | 2004 Aug | 2004 Sep | 2004 Oct | 2004 Nov | 2004 Dec | 2005 Jan | 2005 Jun | 2005 Jul | 2005 Aug | 2005 Oct | 2005 Nov

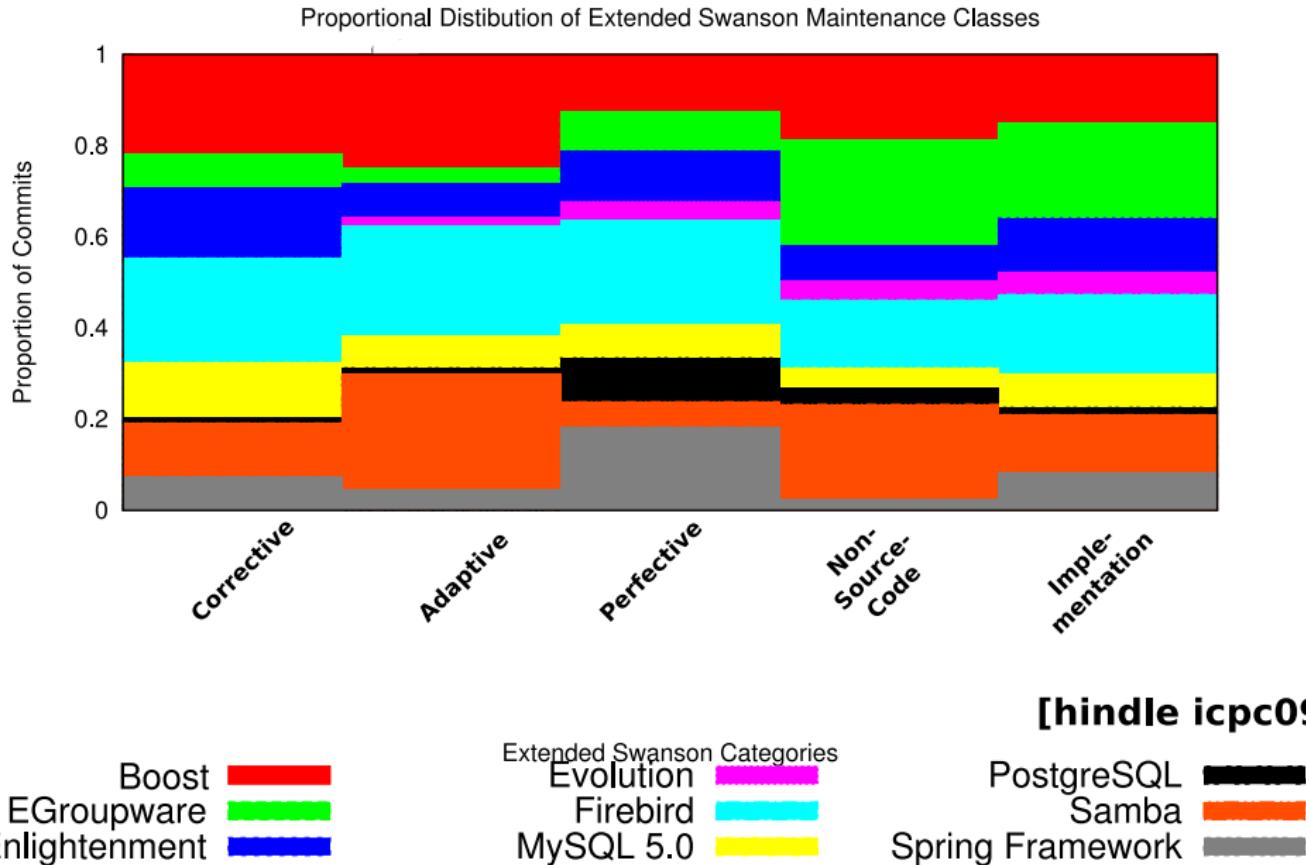
Maintenance Classification



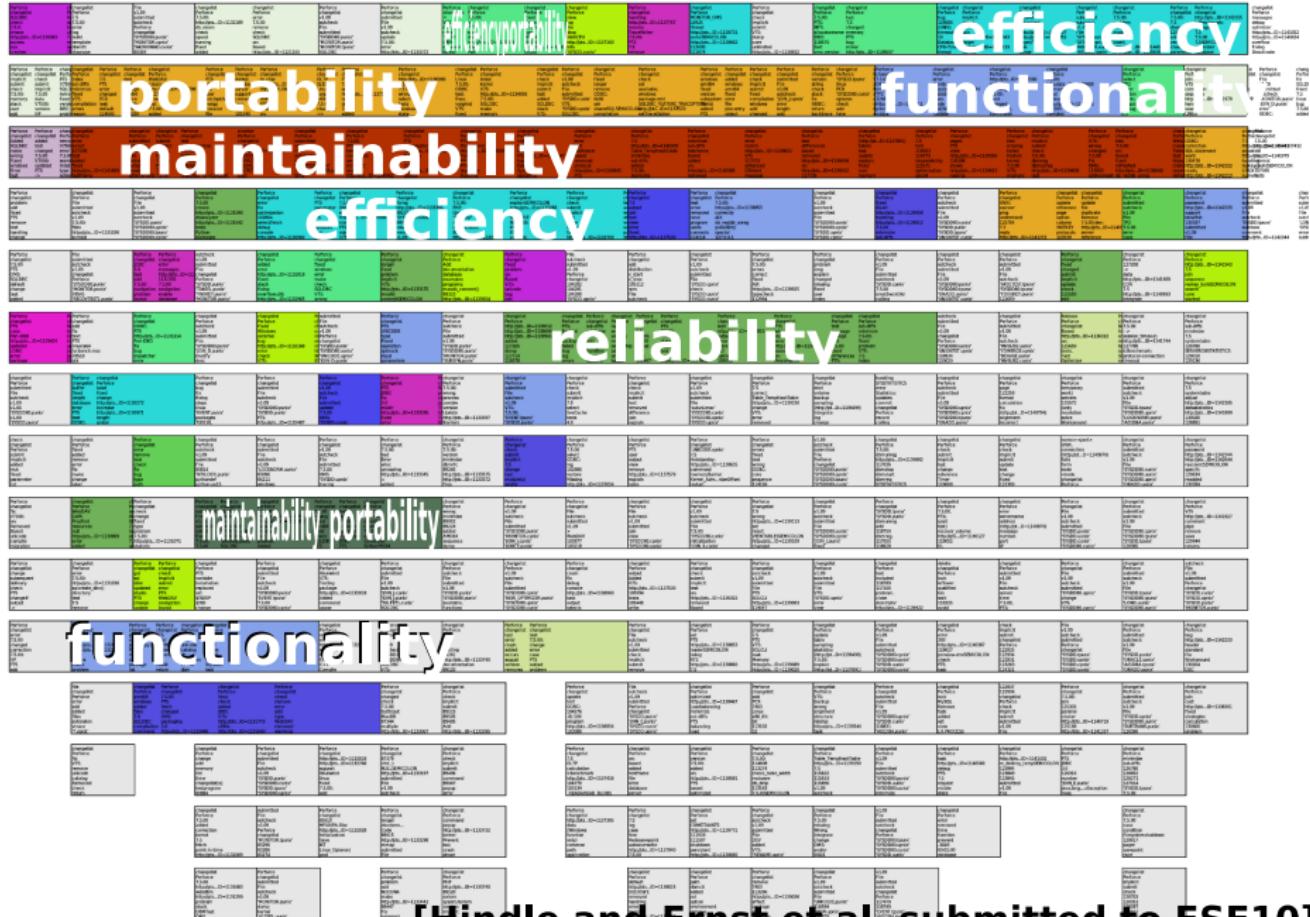
Supervised: Maintenance Classes



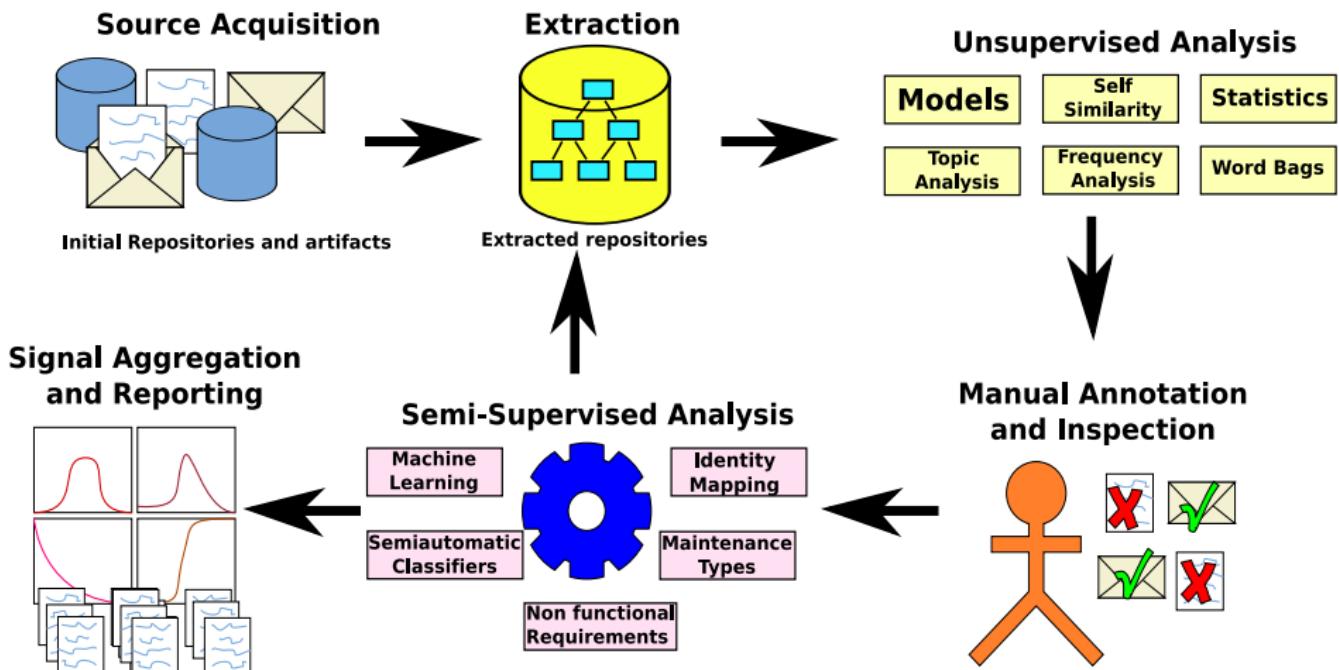
Supervised: Maintenance Classes



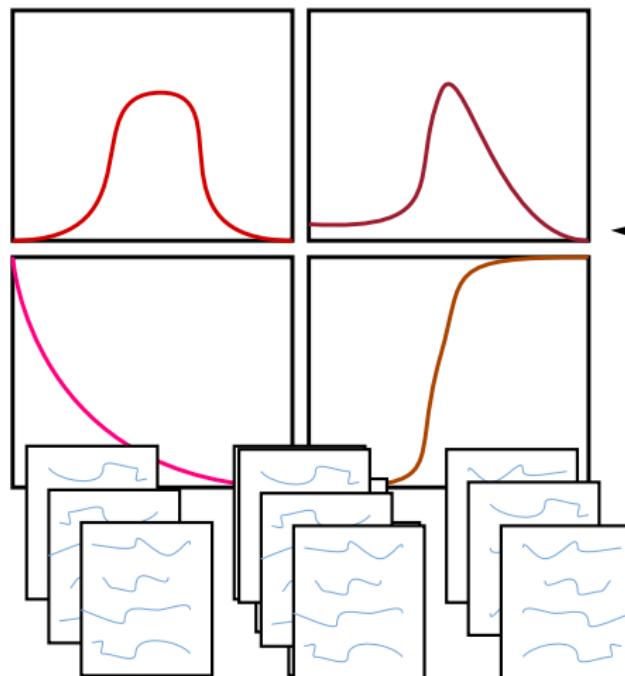
2004 Jun 2004 Jul 2004 Aug 2004 Sep 2004 Oct 2004 Nov 2004 Dec 2005 Jan 2005 Jun 2005 Jul 2005 Aug 2005 Oct 2005 Nov 2005 Dec 2006 Jan 2006 Feb 2006 Mar 2006 Apr 2006 May 2006 Jun



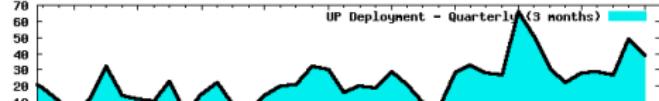
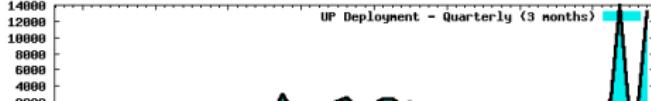
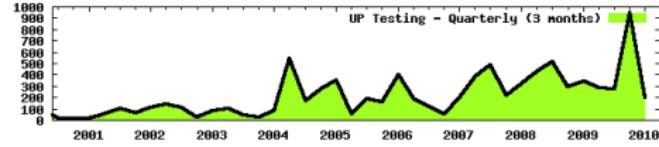
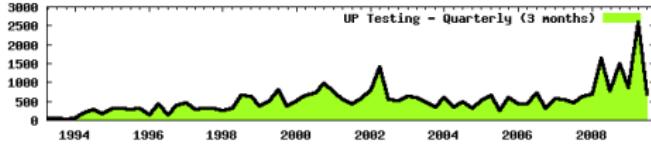
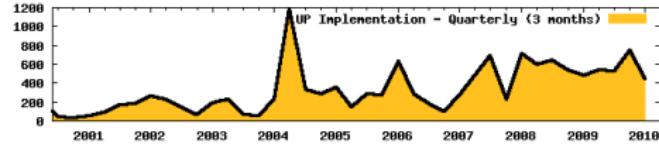
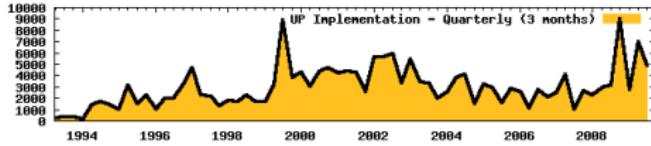
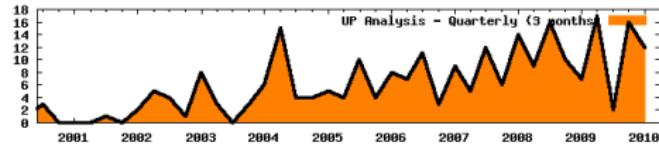
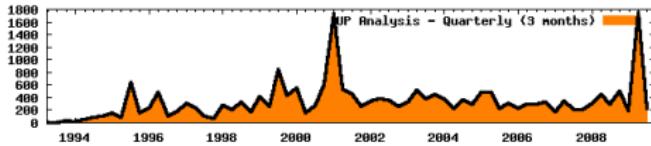
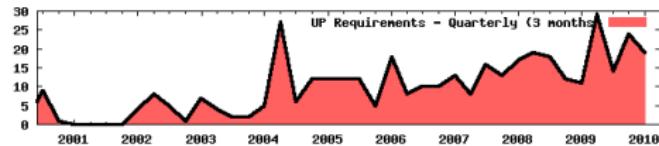
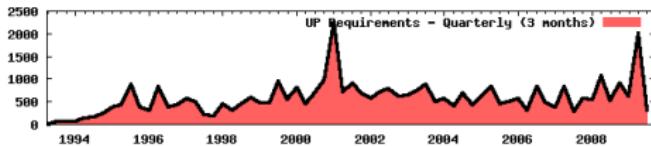
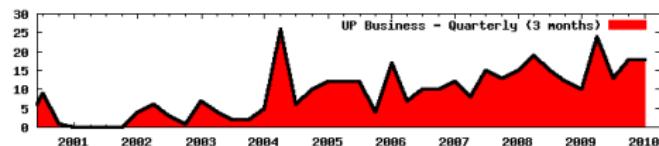
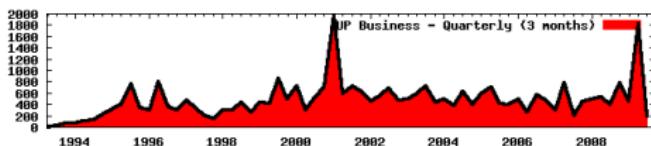
Methodology: Recovered Unified Process Views



Signal Aggregation and Reporting



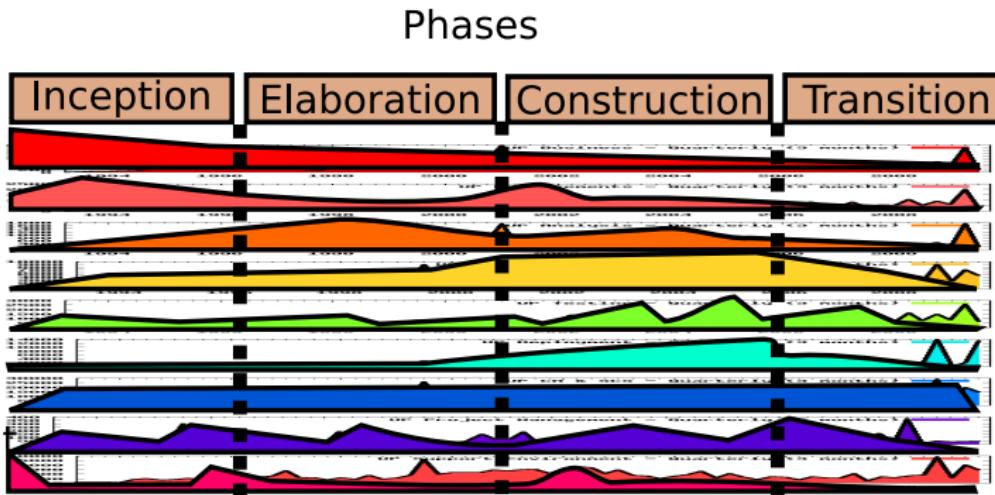
Reporting



Recovered Unified Process Views

Disciplines

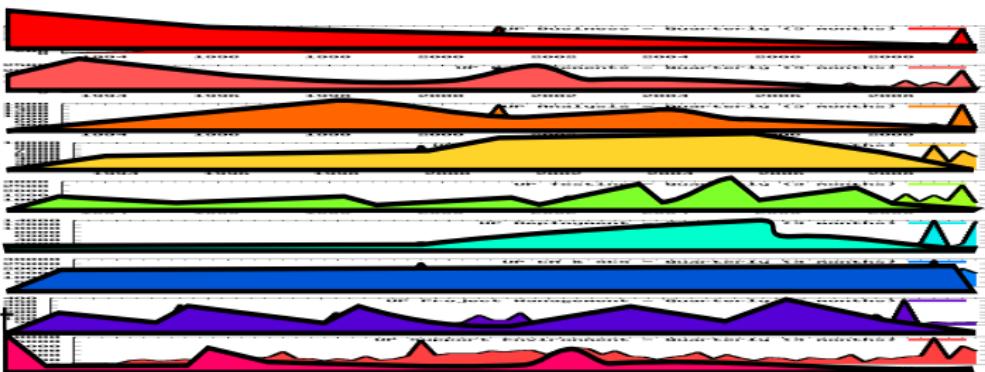
Business Modeling
Requirements
Analysis & Design
Implementation
Test
Deployment
CM and SCS
Project Management
Environment



Recovered Unified Process Views

Disciplines

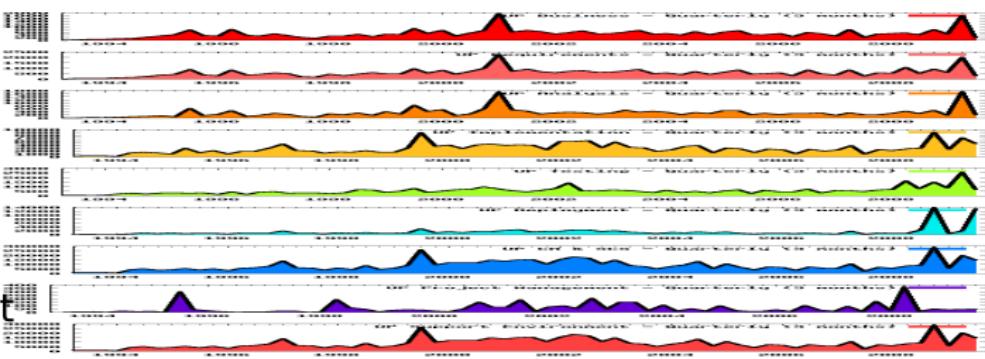
Business Modeling
Requirements
Analysis & Design
Implementation
Test
Deployment
CM and SCS
Project Management
Environment



Recovered Unified Process Views

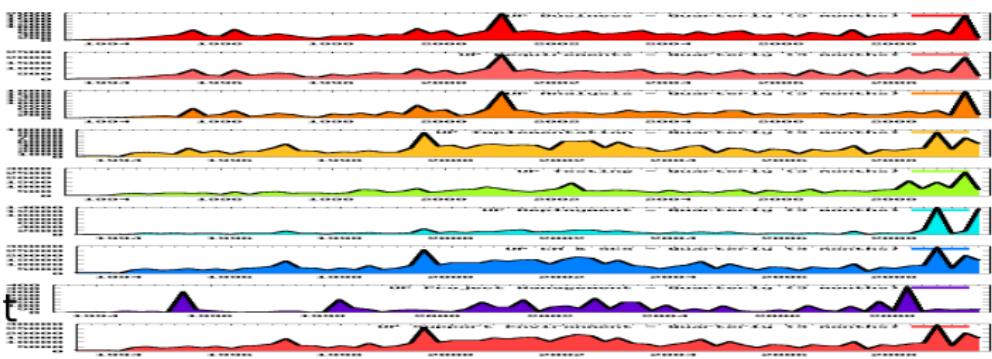
Disciplines

Business Modeling
Requirements
Analysis & Design
Implementation
Test
Deployment
CM and SCS
Project Management
Environment



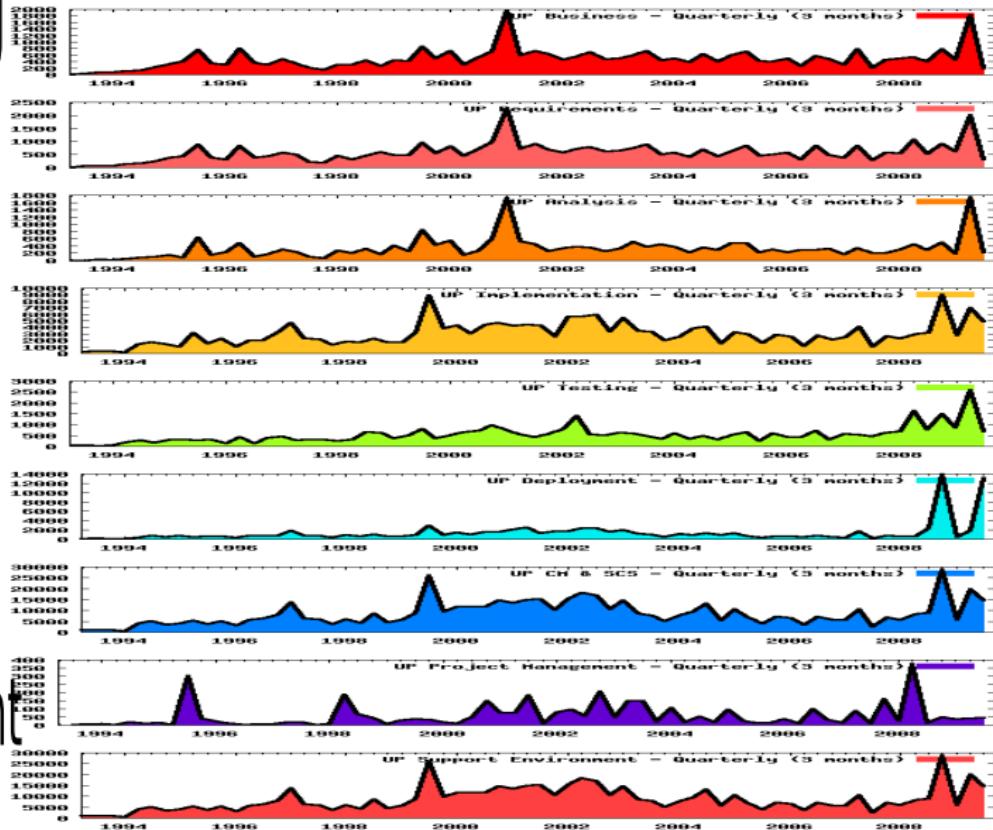
Recovered Unified Process Views

Business Modeling
Requirements
Analysis & Design
Implementation
Test
Deployment
CM and SCS
Project Management
Environment



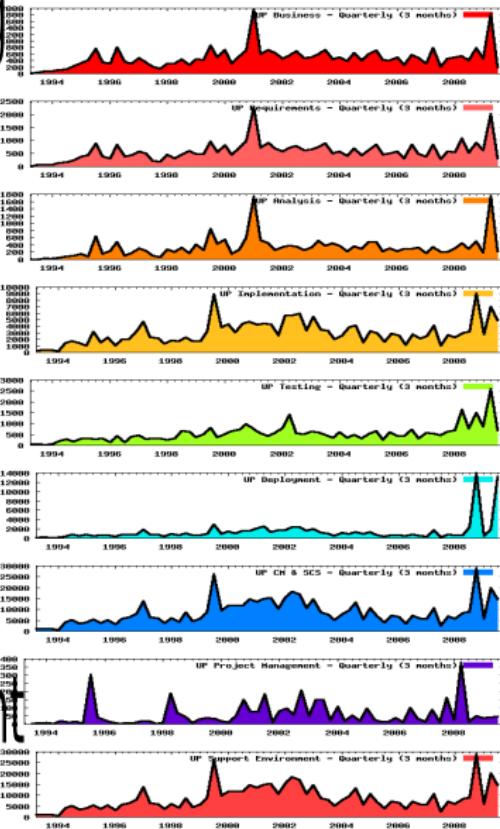
Recovered Unified Process Views

Business Modeling
Requirements
Analysis & Design
Implementation
Test
Deployment
CM and SCS
Project Management
Environment



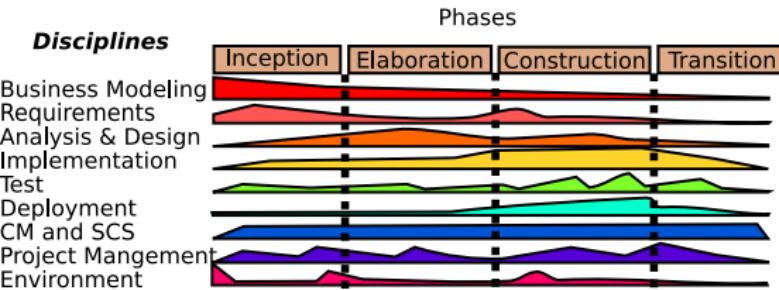
Recovered Unified Process Views

Business Modeling
Requirements
Analysis & Design
Implementation
Test
Deployment
CM and SCS
Project Management
Environment

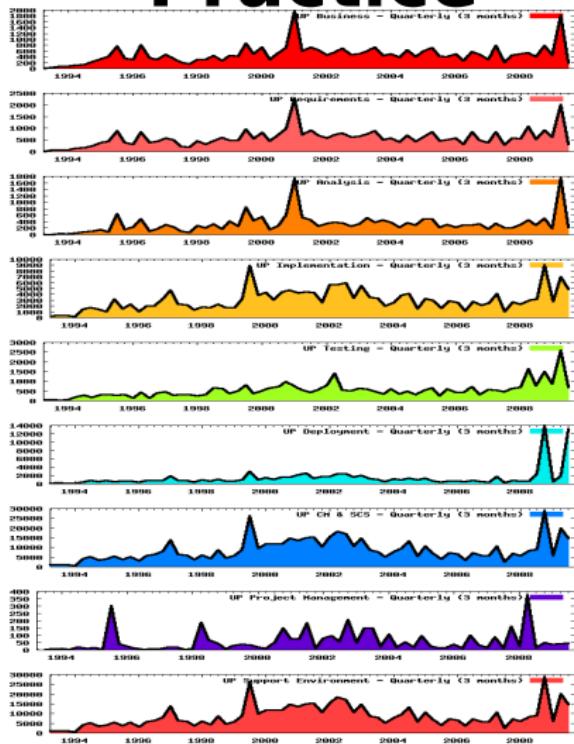


Recovered Unified Process Views

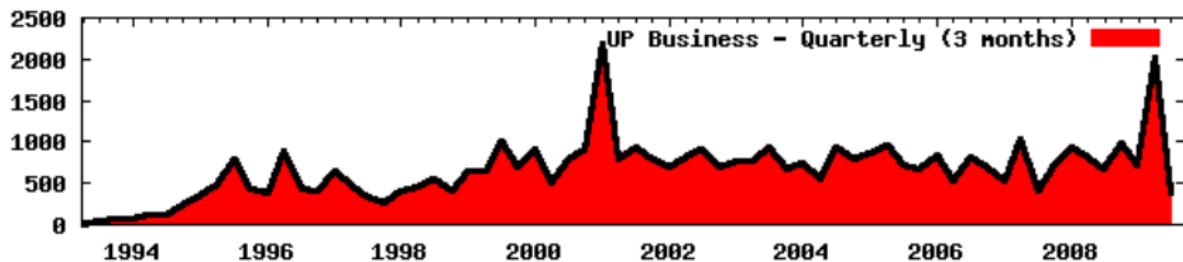
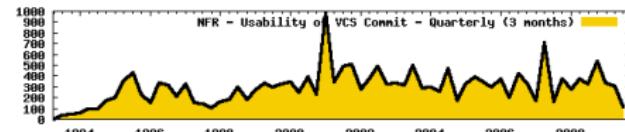
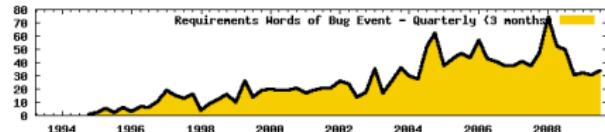
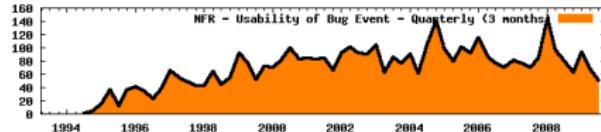
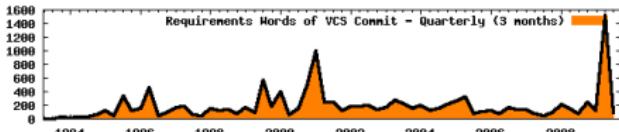
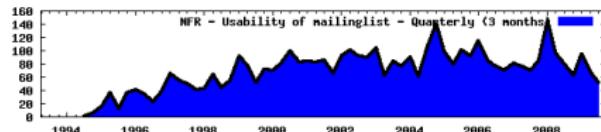
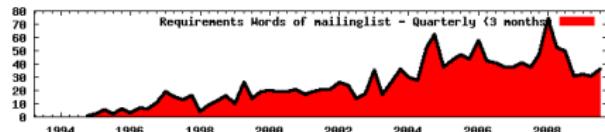
Theory



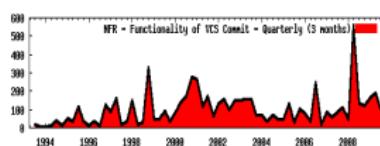
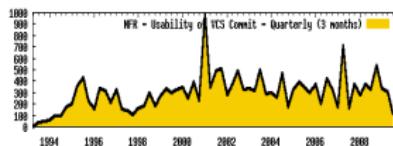
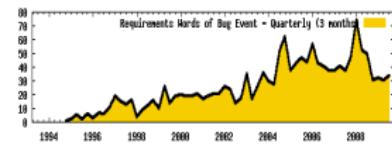
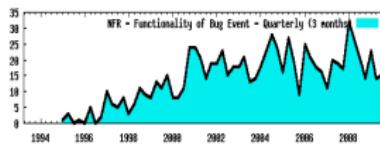
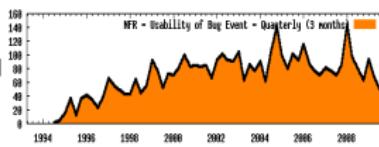
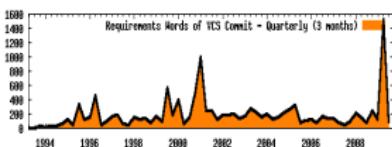
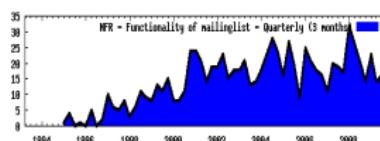
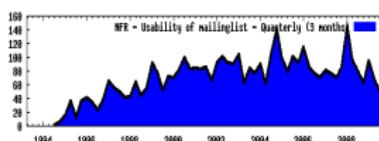
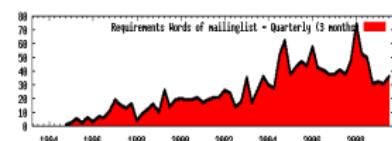
Practice



UP Business Modelling Signal



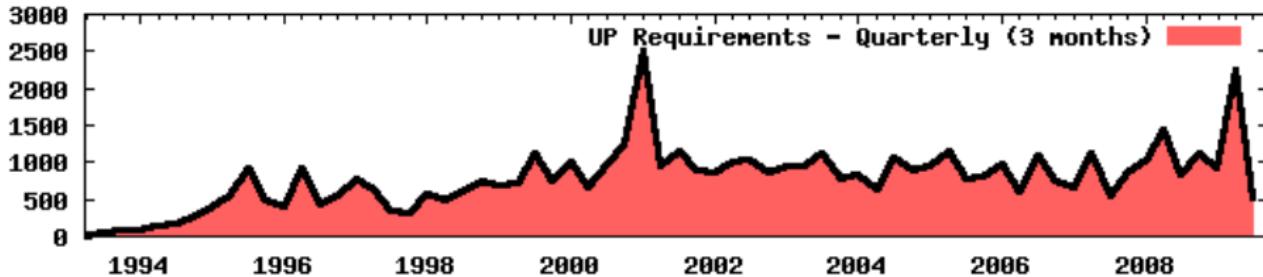
UP Requirements Signal



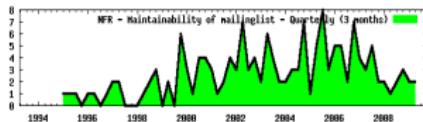
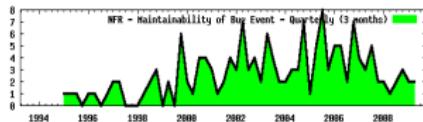
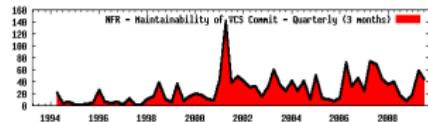
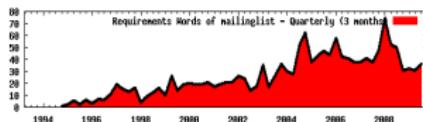
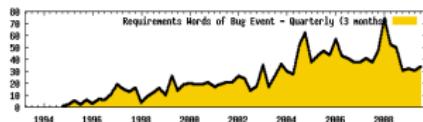
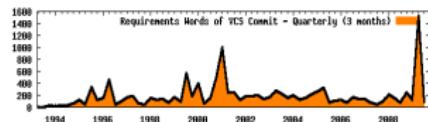
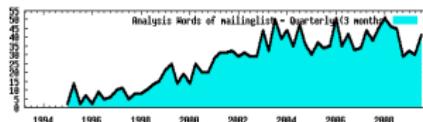
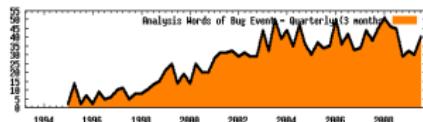
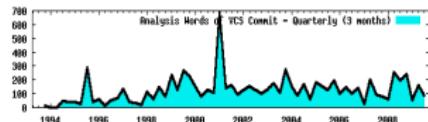
+

+

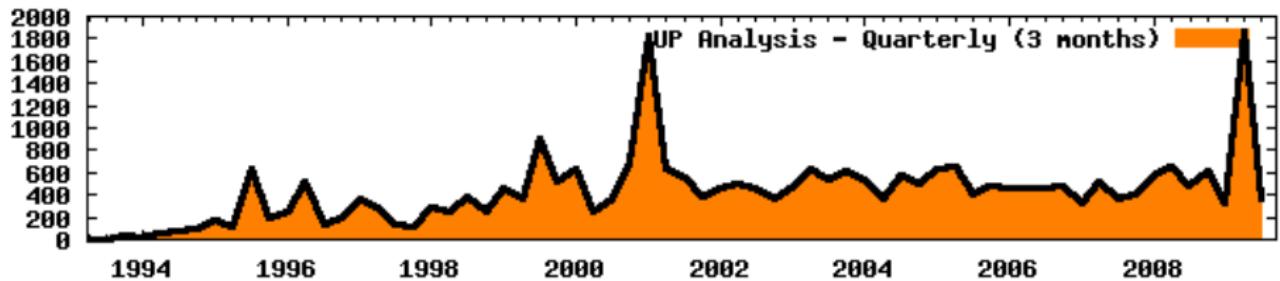
=



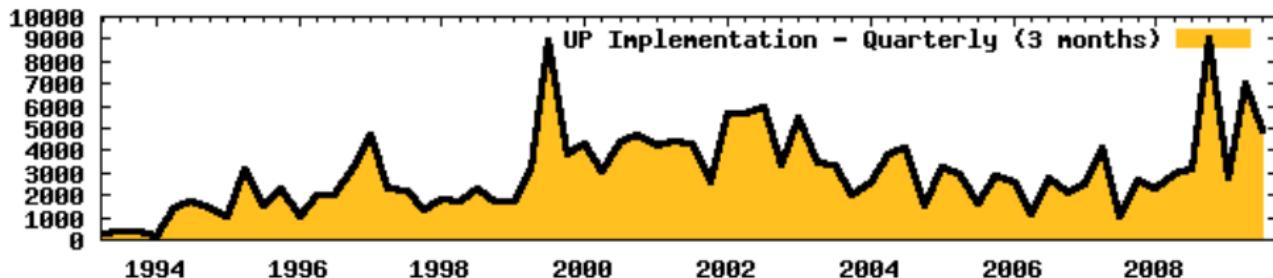
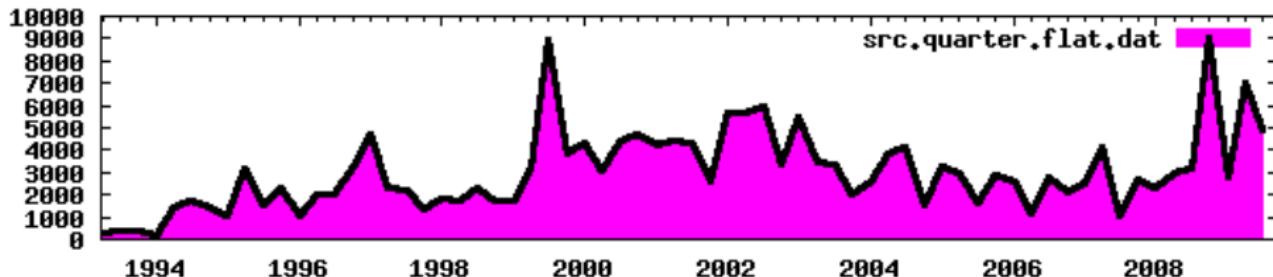
UP Analysis Signal



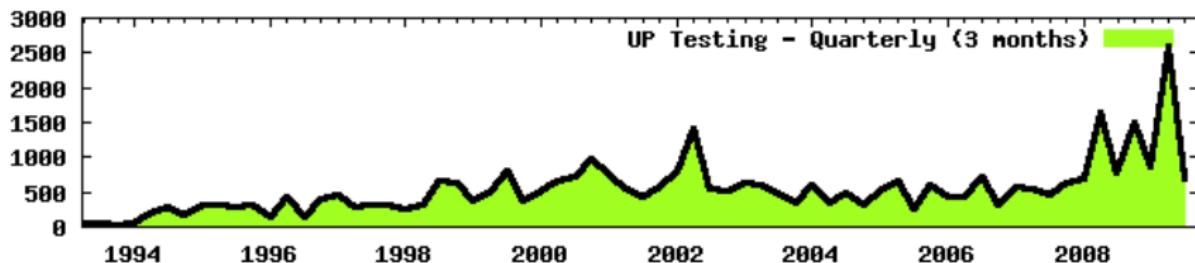
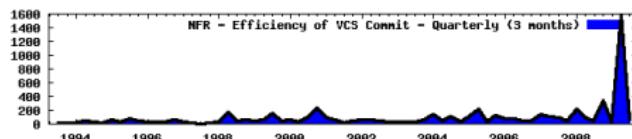
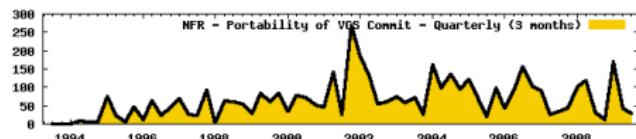
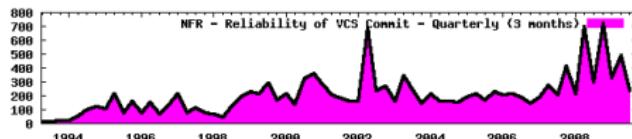
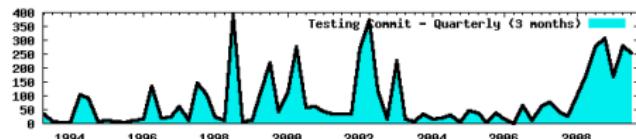
====



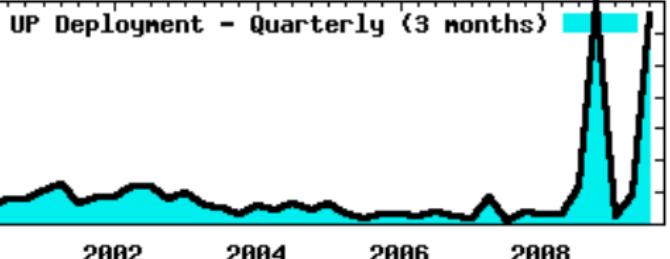
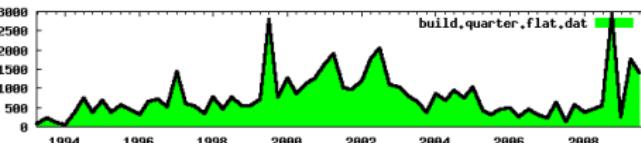
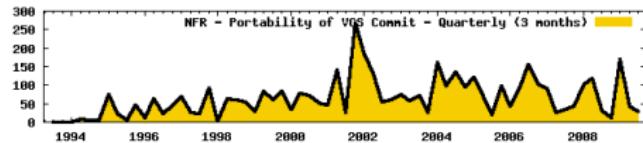
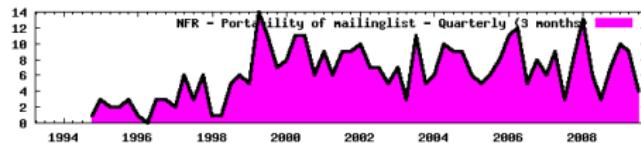
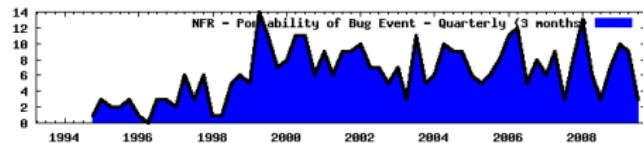
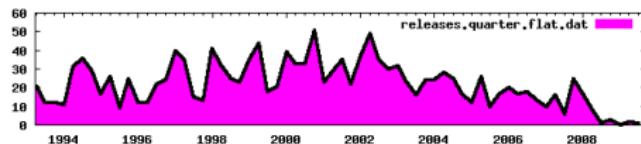
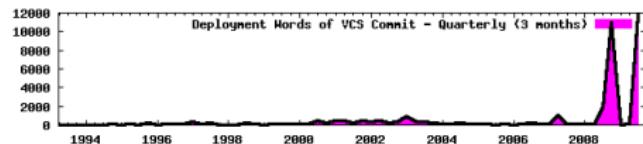
UP Implementation Signal



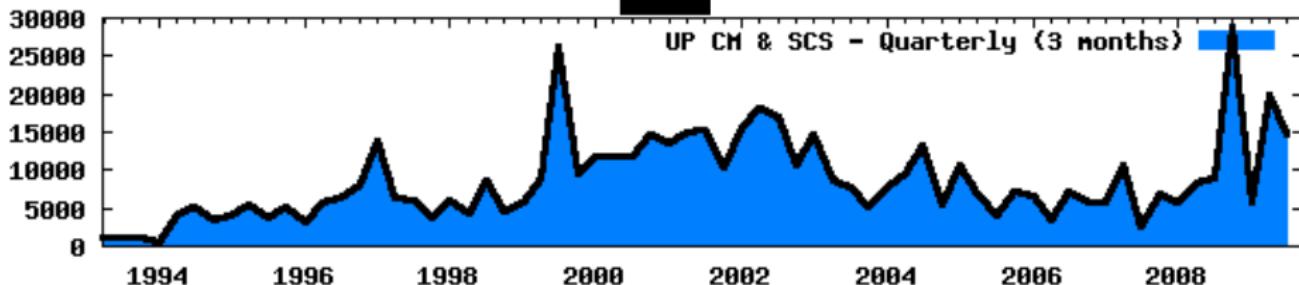
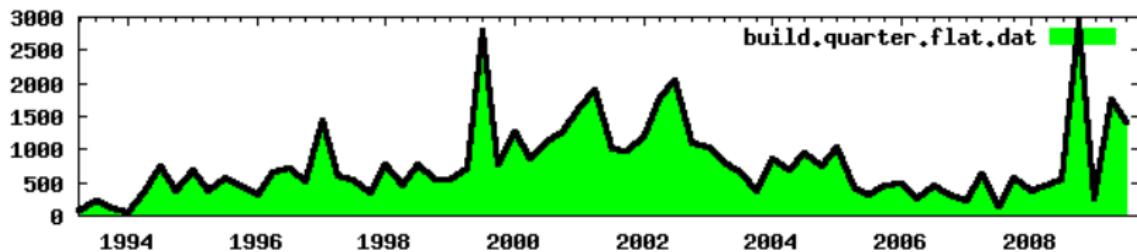
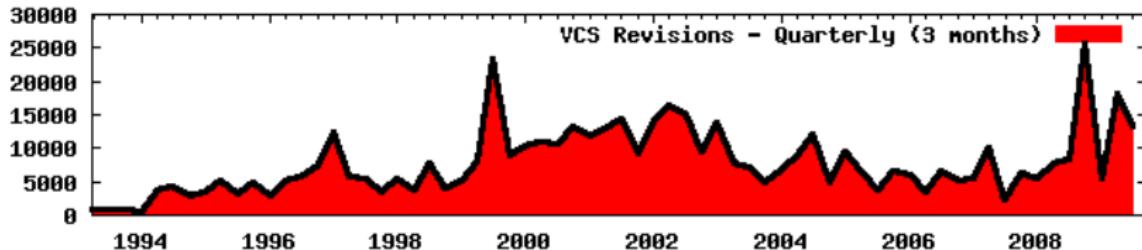
UP Testing Signal



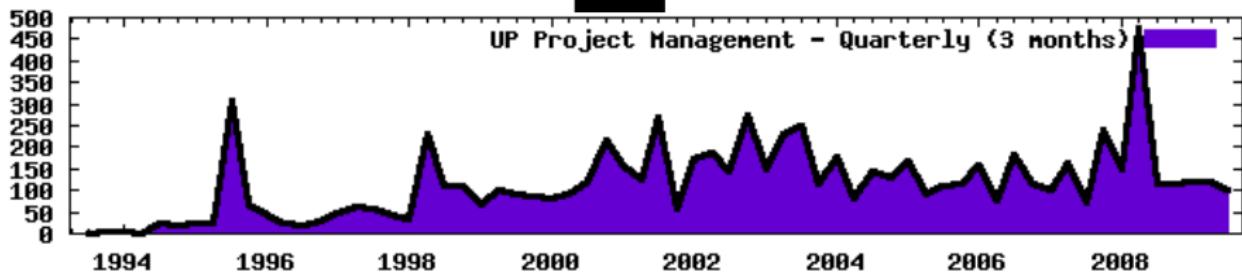
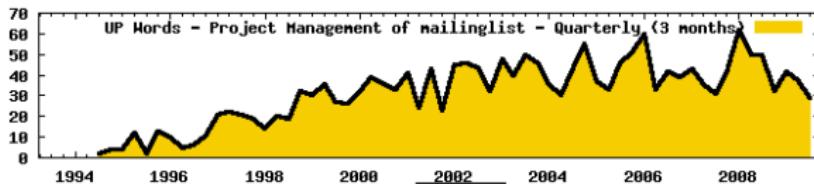
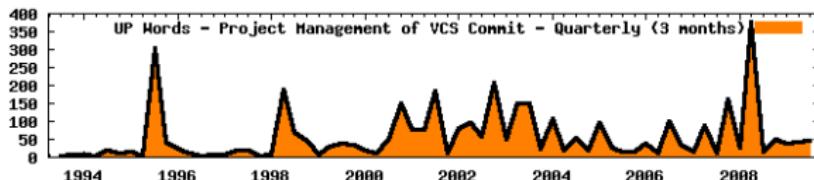
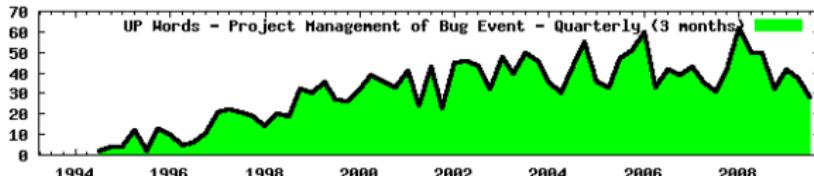
UP Deployment Signal



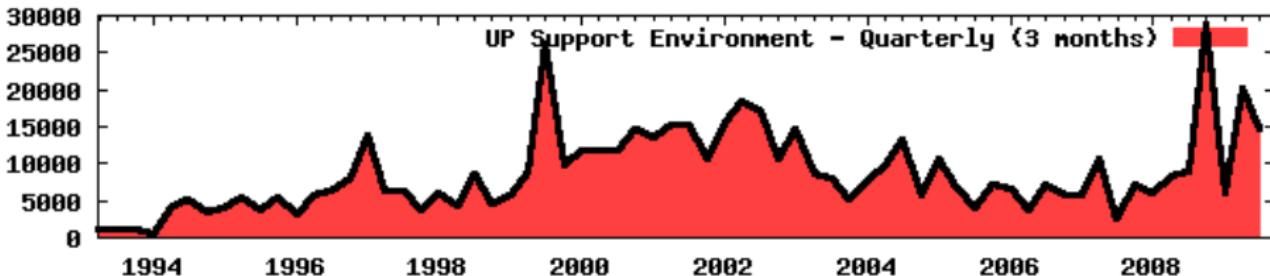
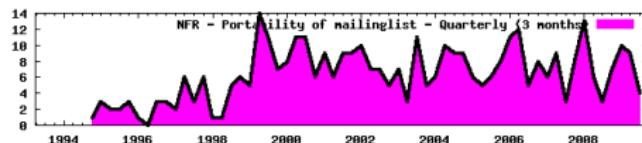
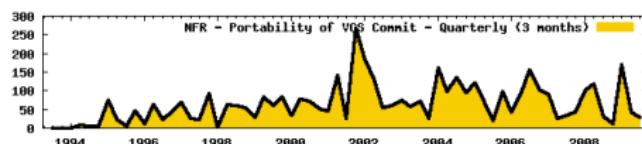
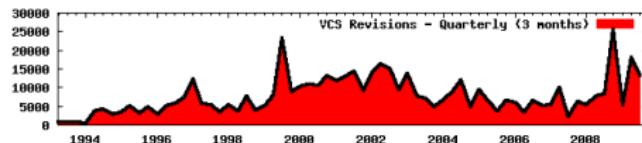
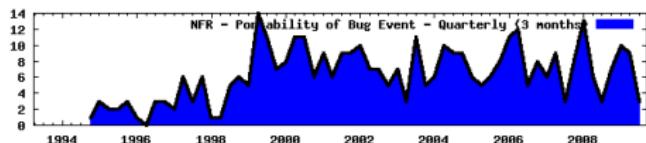
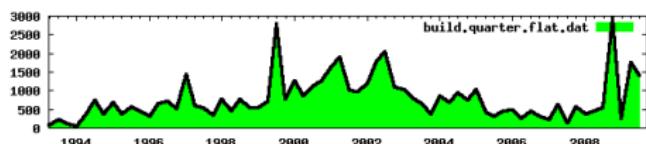
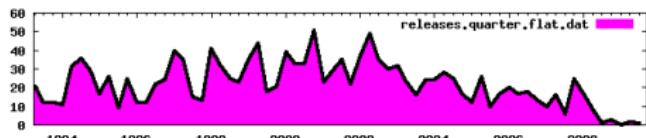
UP Configuration Management and SCS



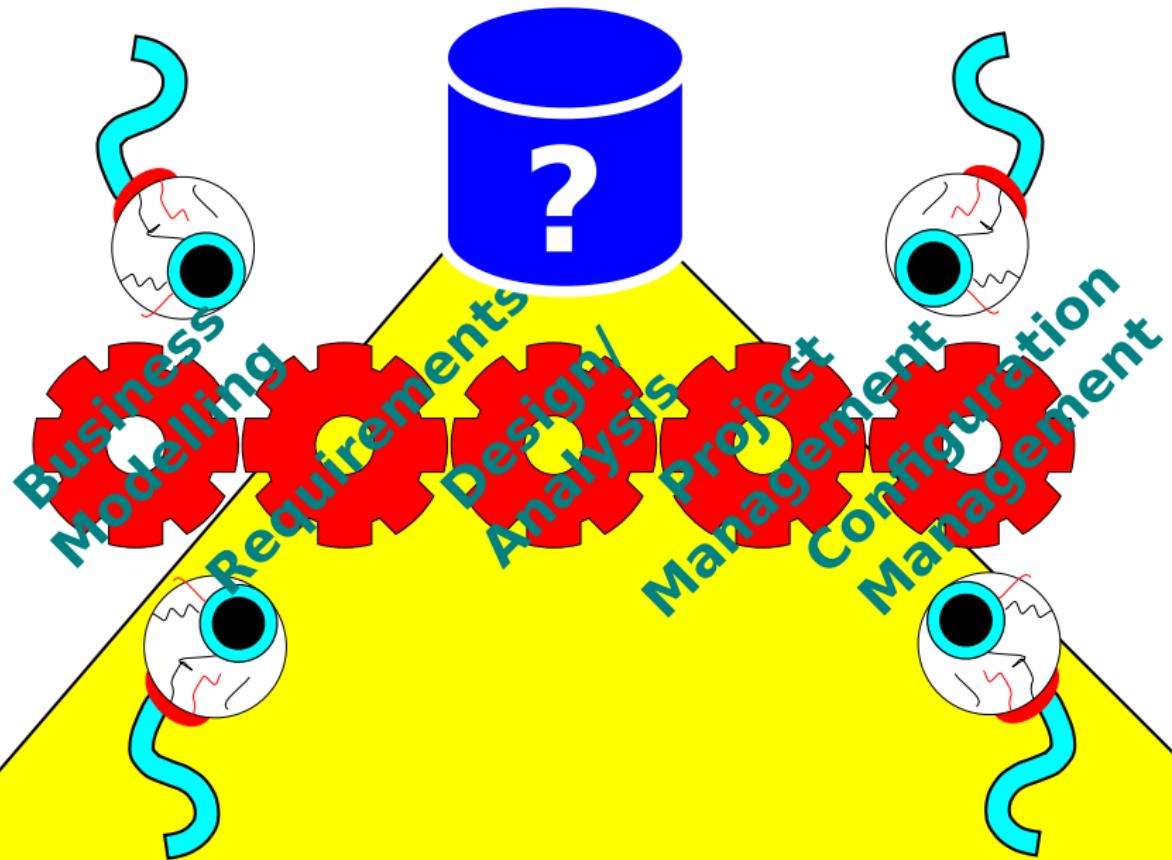
UP Project Management Signal



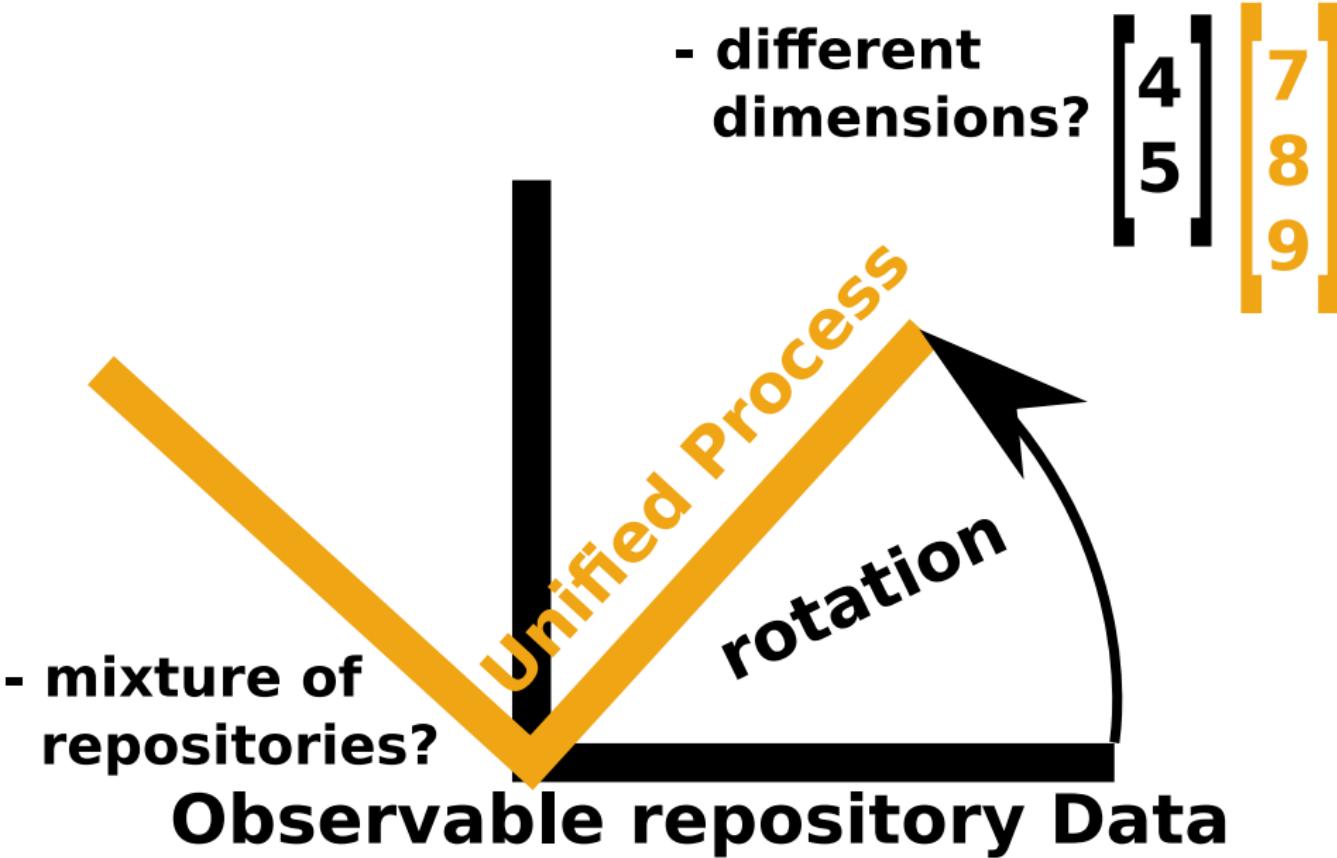
UP Environment Signal



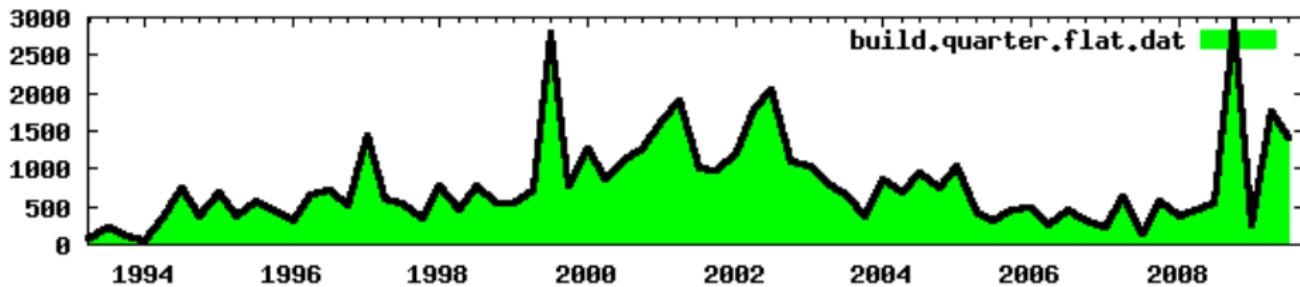
UP Observability



UP Observability



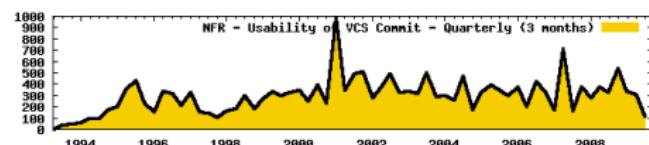
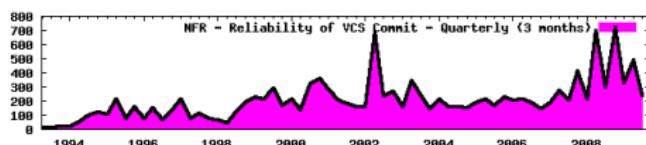
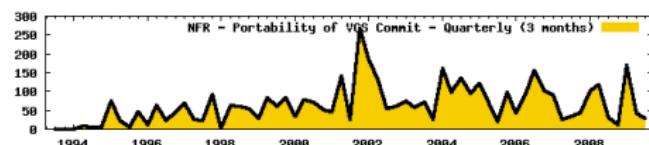
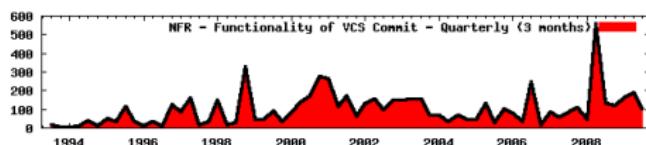
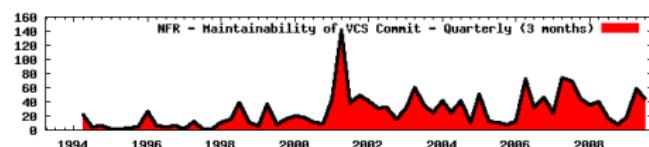
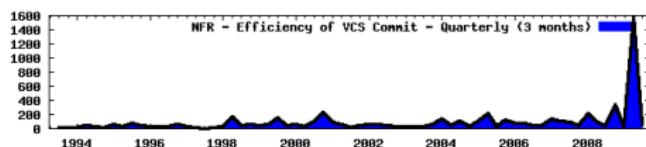
Process Heavy Signals: Build Commits



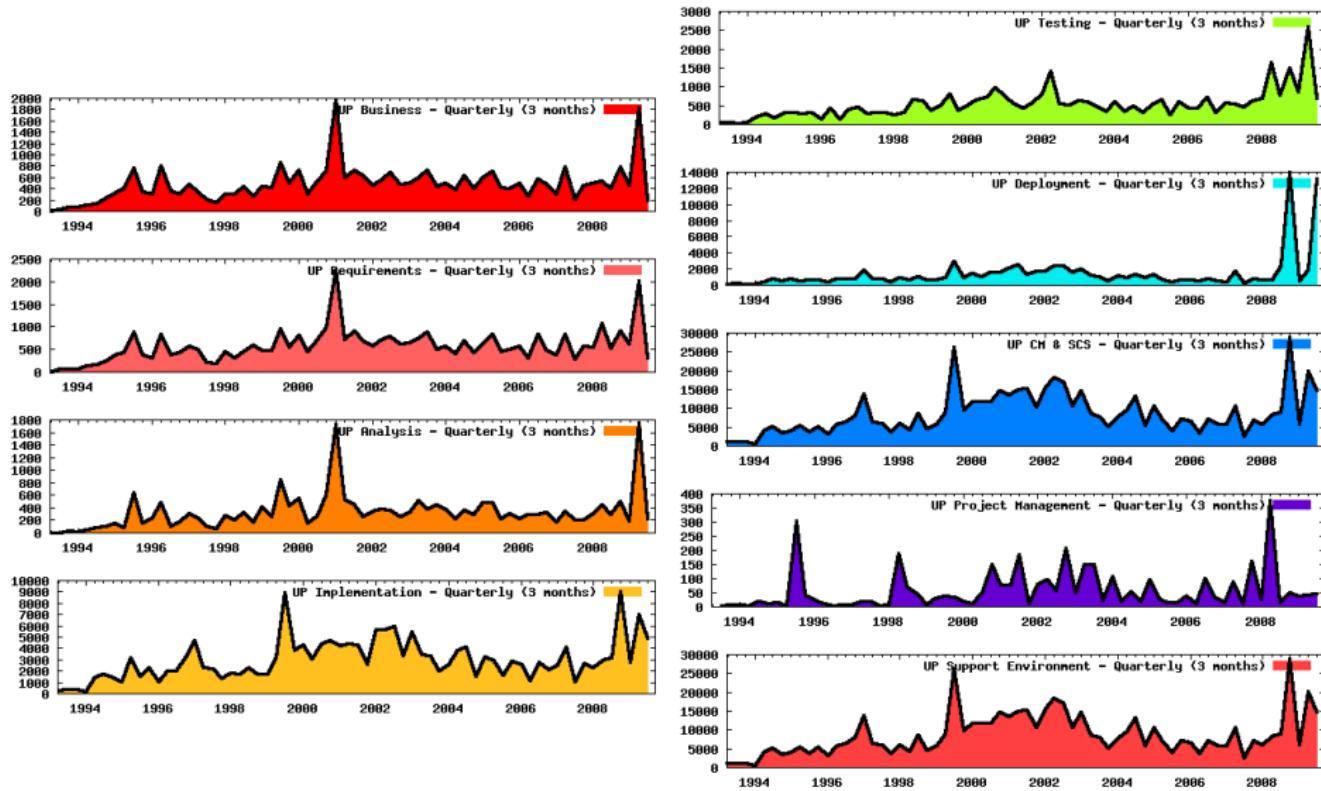
Related to

- portability**
- change in modularity**
- feature addition and removal**

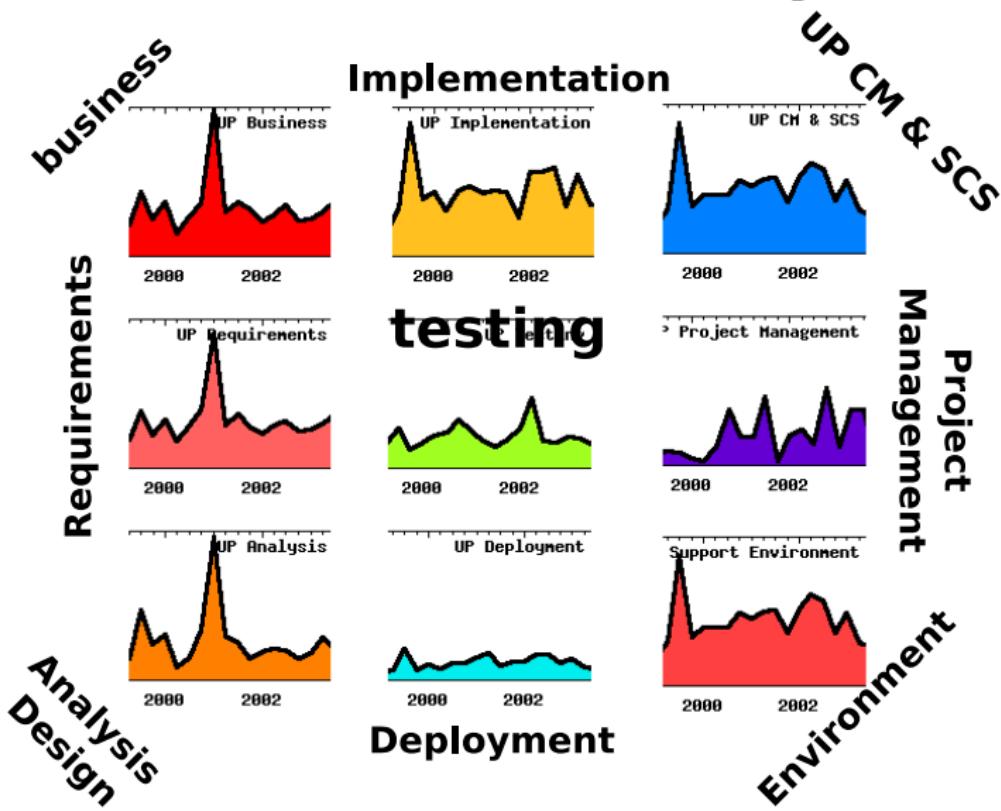
Process Heavy Signals: Non Functional Requirements



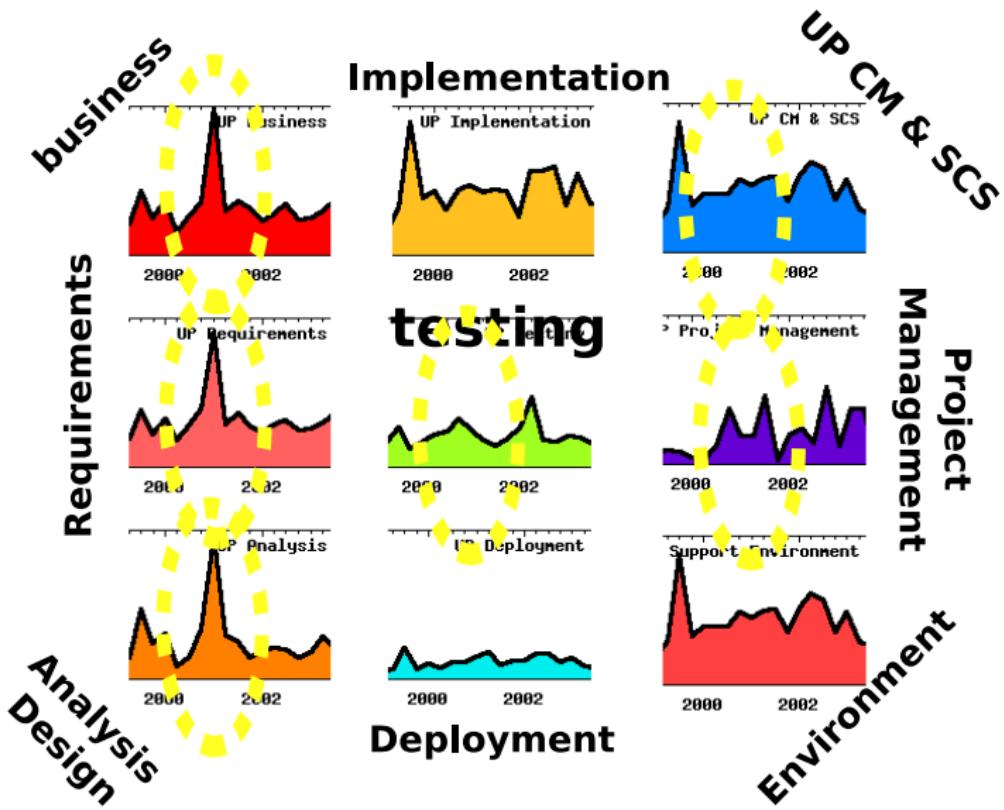
FreeBSD Case Study



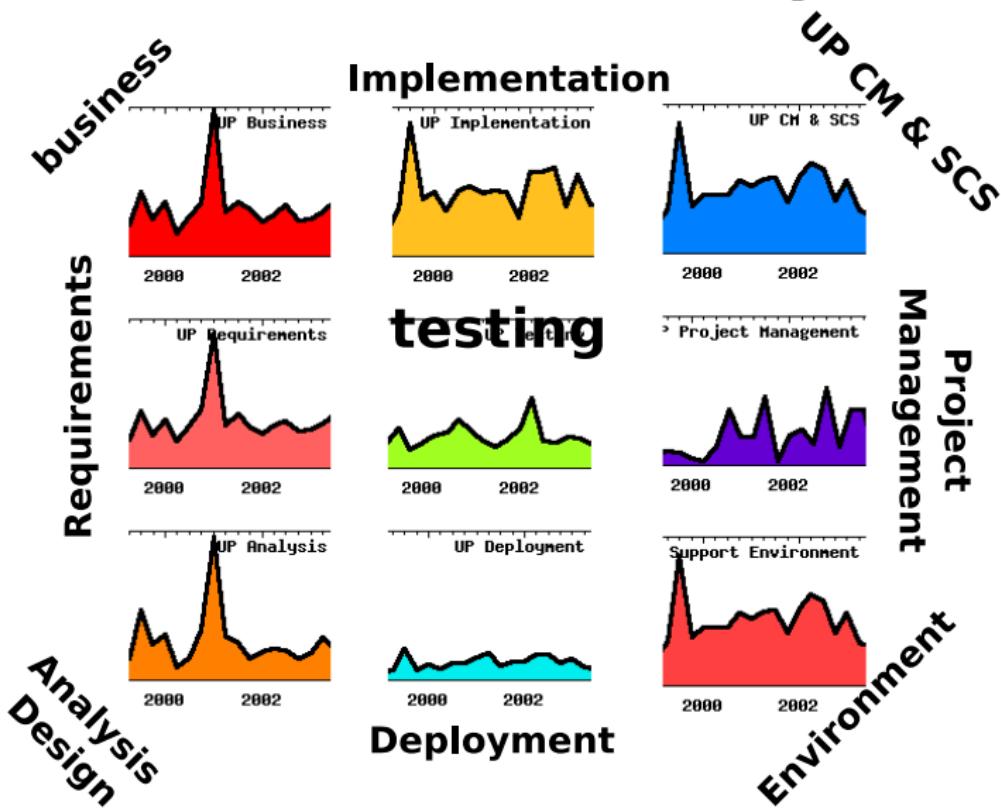
FreeBSD Case Study: 2001



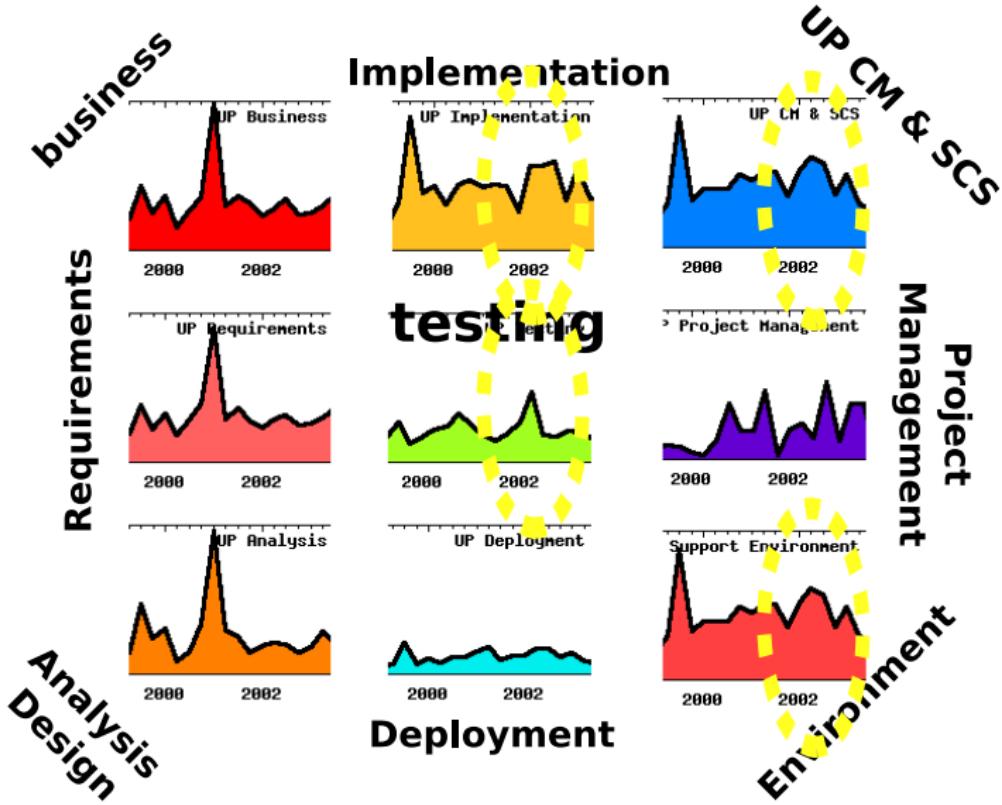
FreeBSD Case Study: 2001



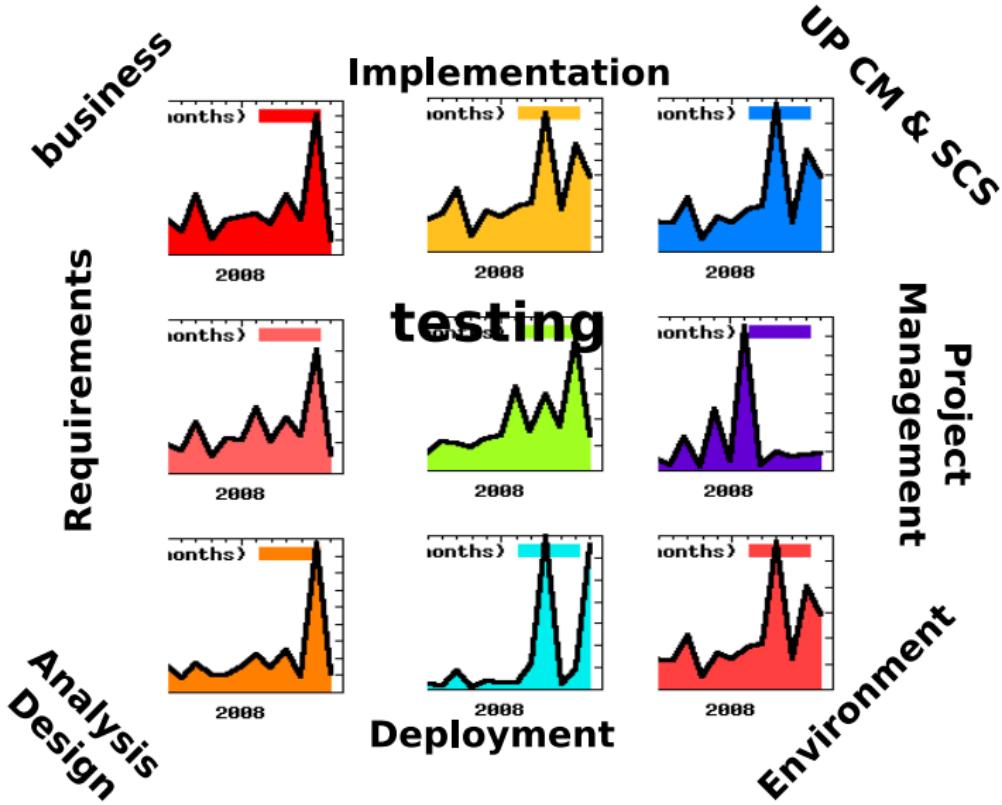
FreeBSD Case Study: 2002



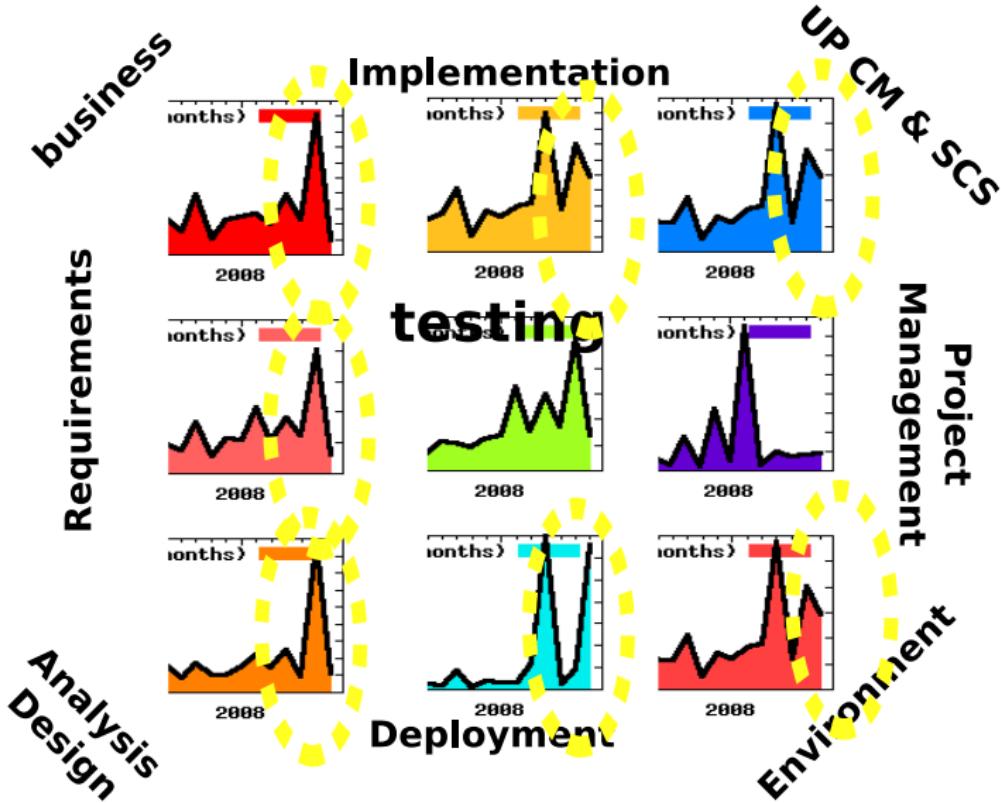
FreeBSD Case Study: 2002



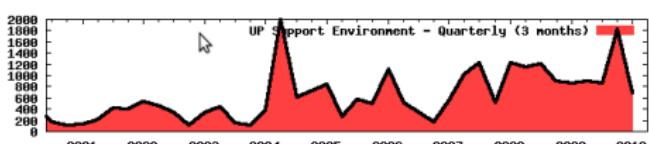
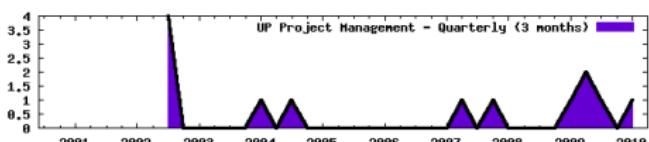
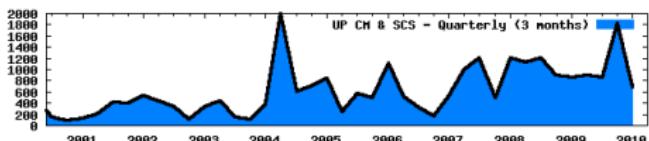
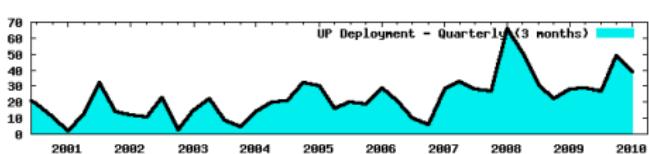
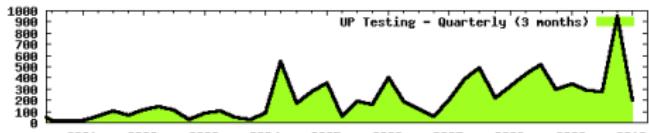
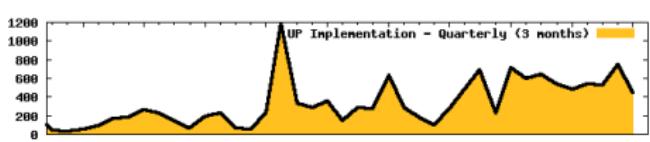
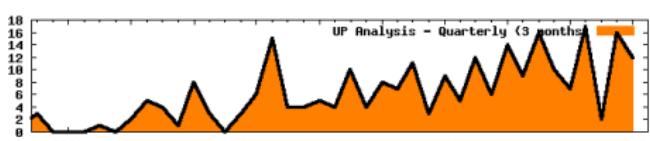
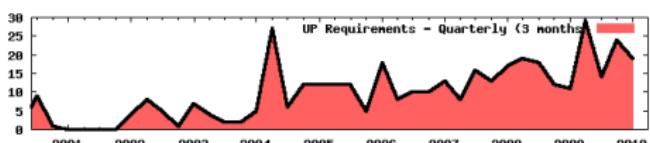
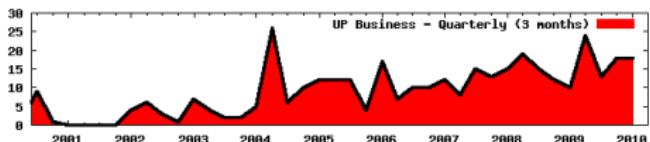
FreeBSD Case Study: 2009



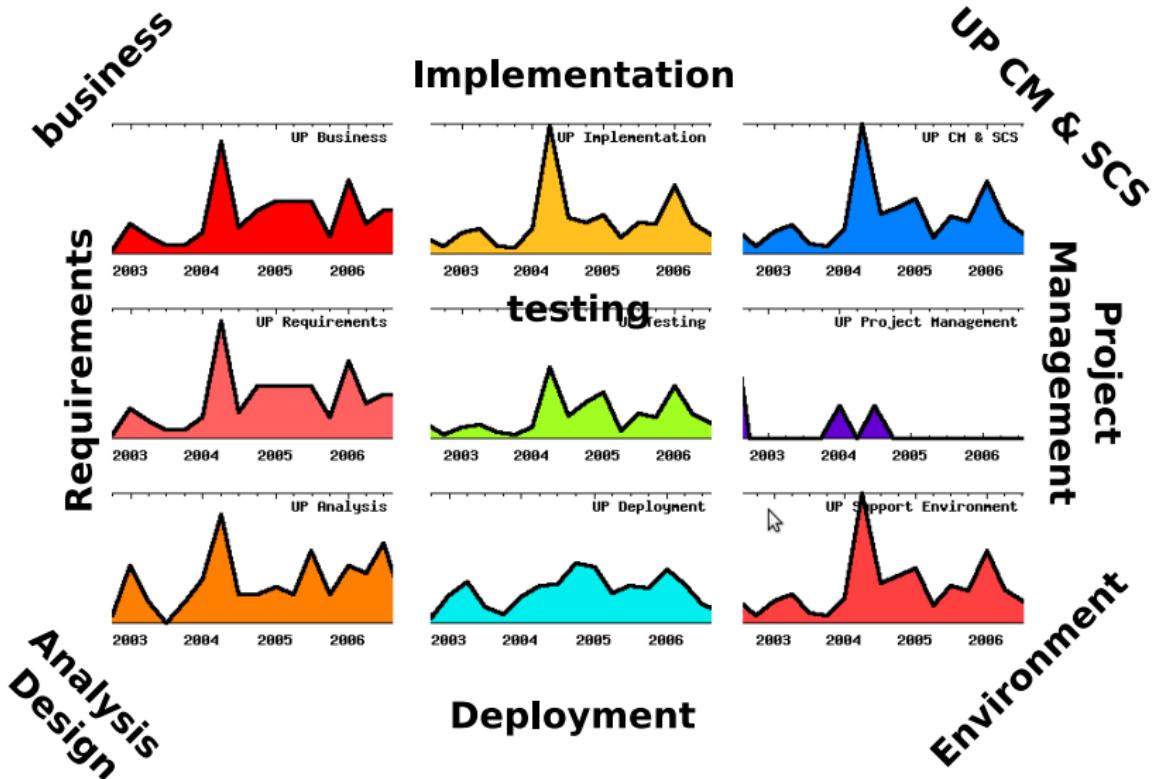
FreeBSD Case Study: 2009



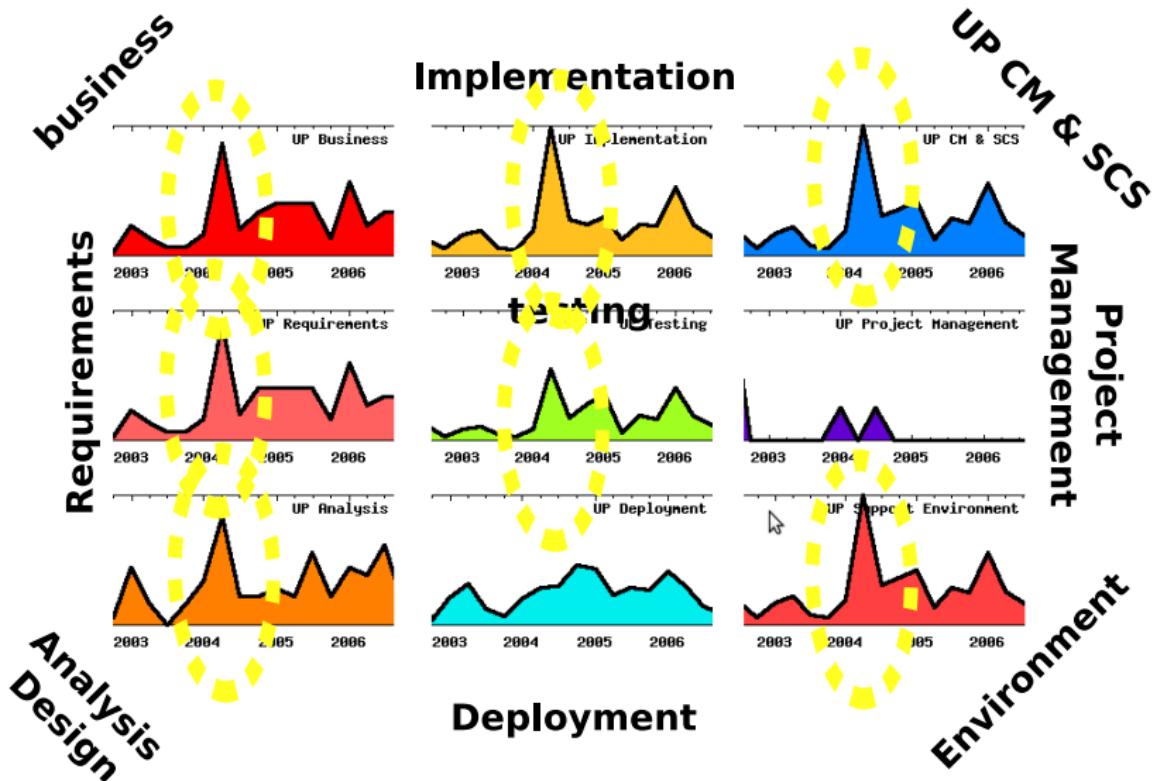
SQLite Case Study



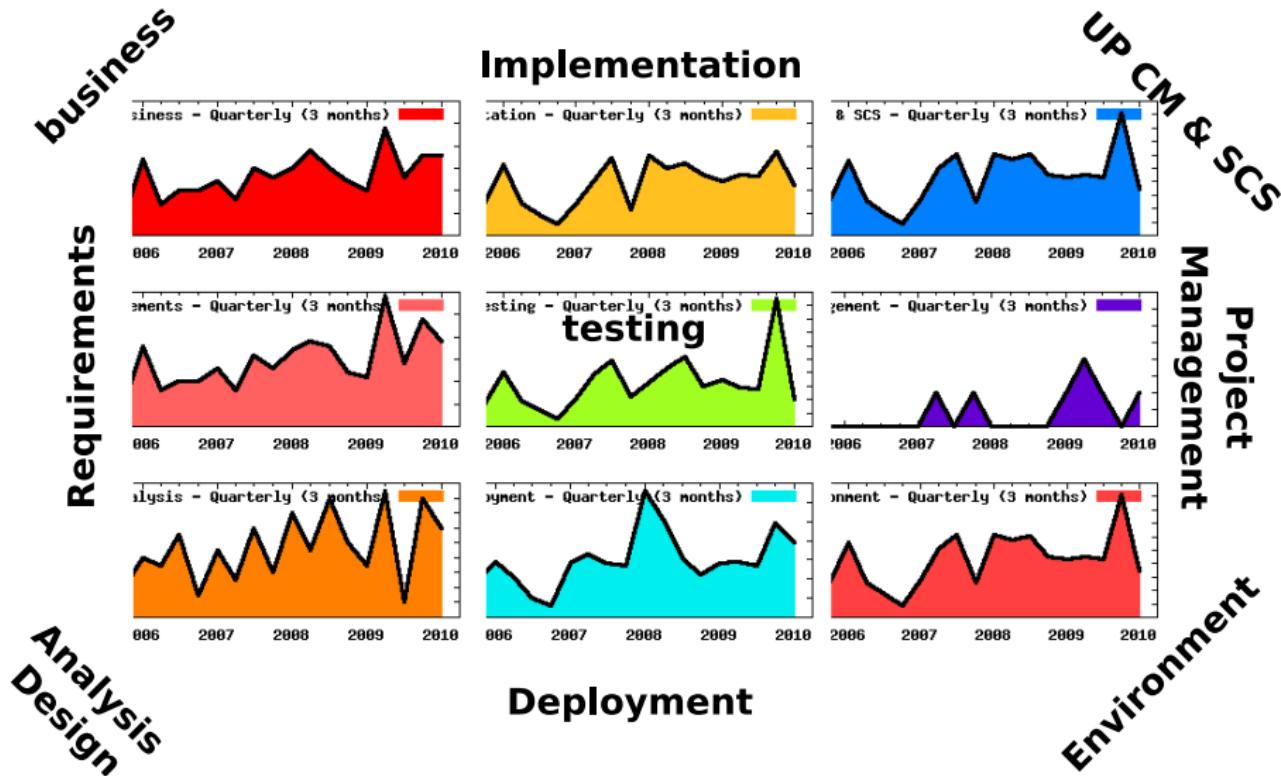
SQLite Case Study: 2004



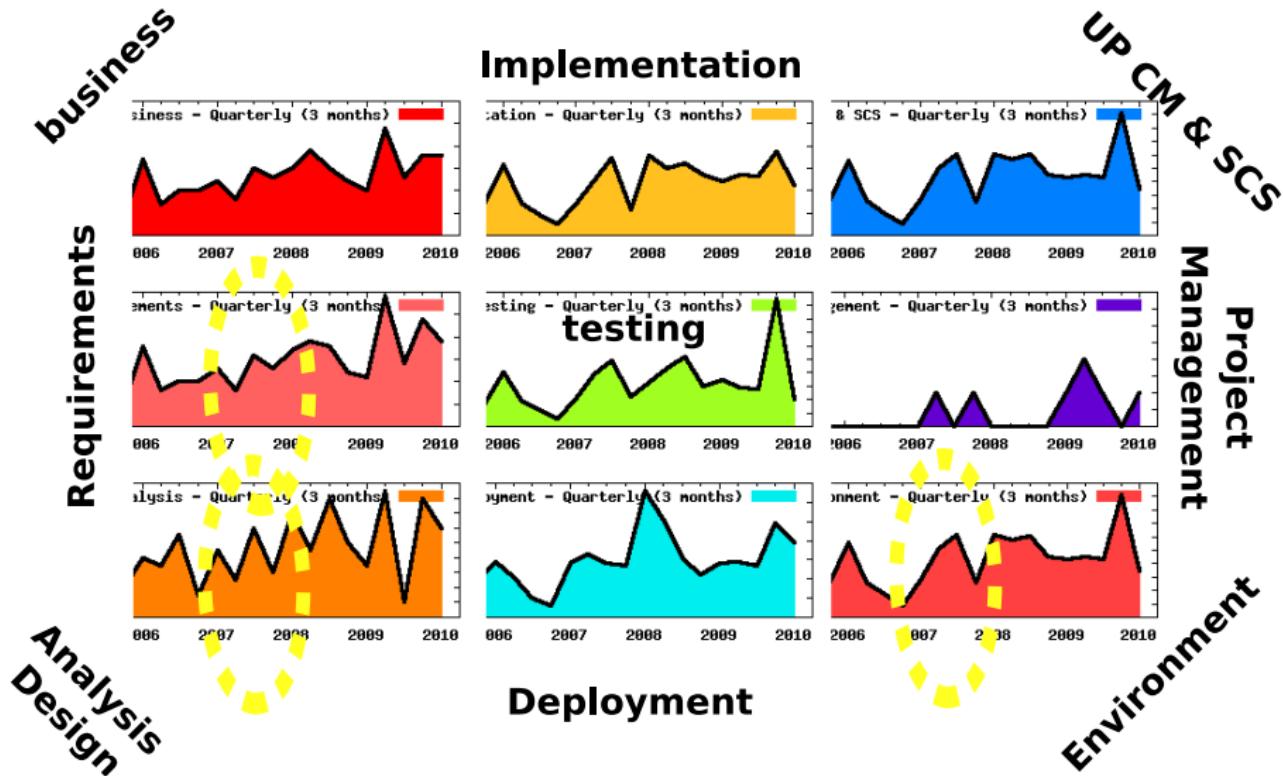
SQLite Case Study: 2004



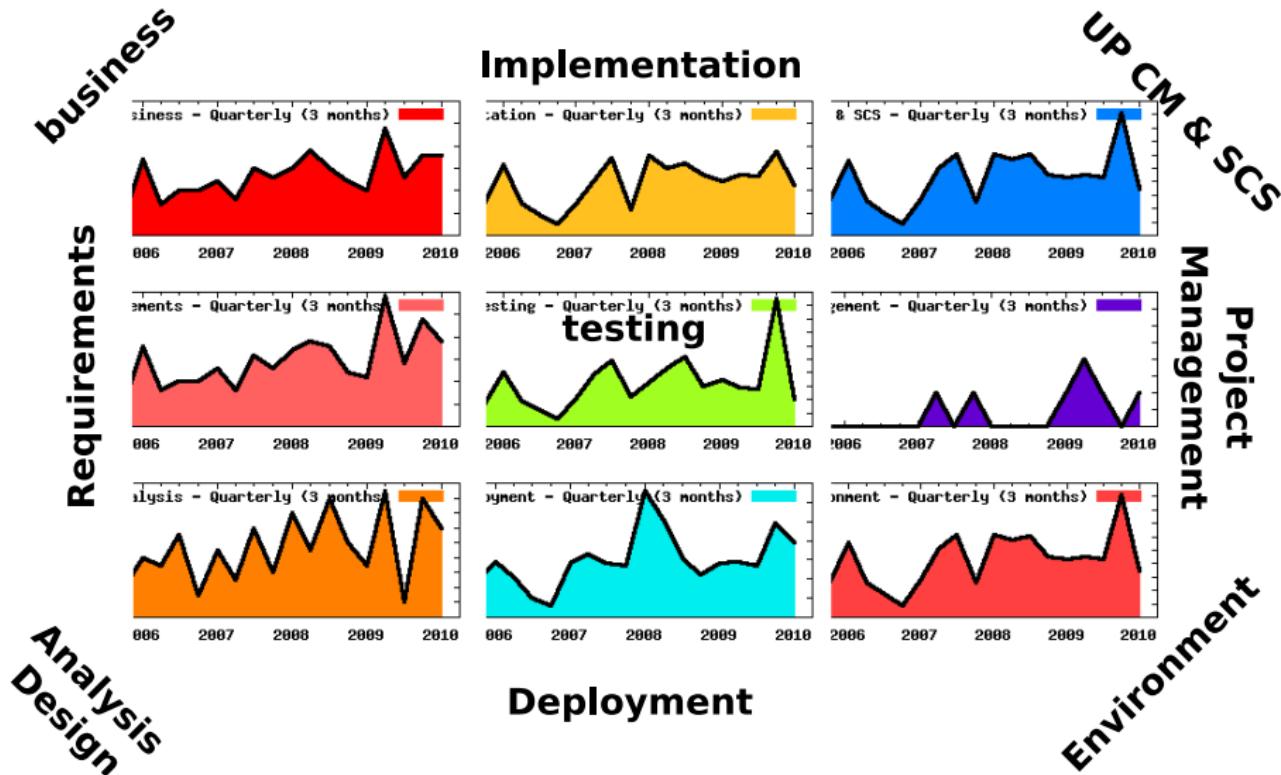
SQLite Case Study: 2007-2008



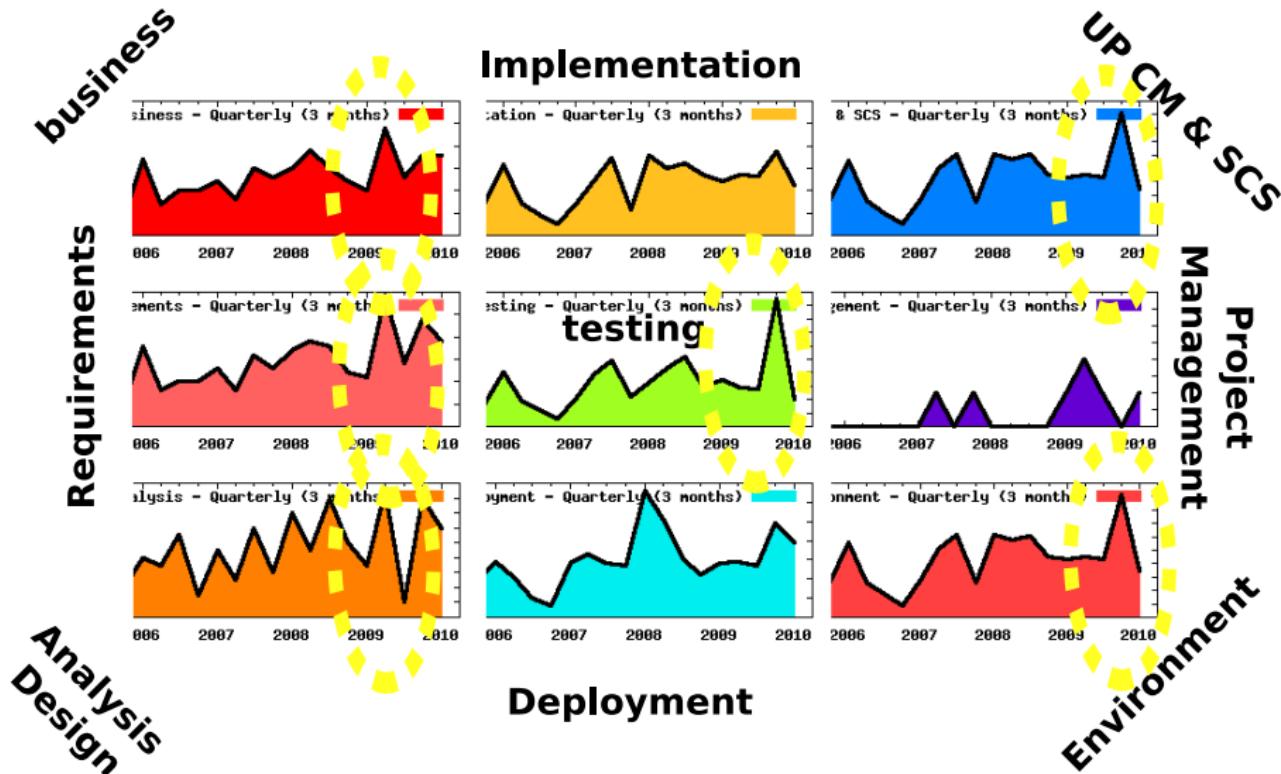
SQLite Case Study: 2007-2008



SQLite Case Study: 2009



SQLite Case Study: 2009



What have we done?

Theory

Business Modeling



Requirements



Analysis & Design



Implementation



Test



Deployment



CM and SCS



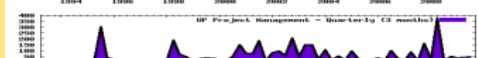
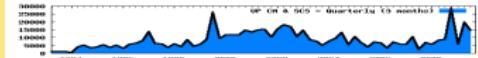
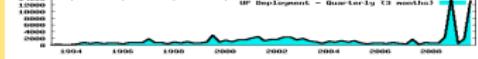
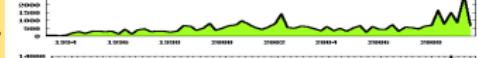
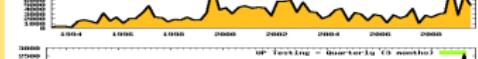
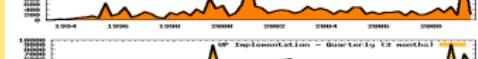
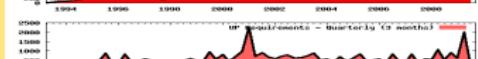
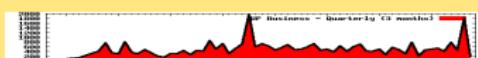
Project Management



Environment

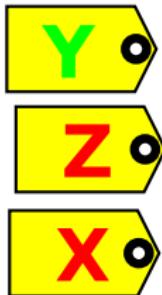


Practice



Looking Forward

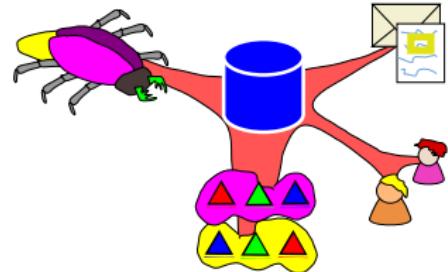
What can other tools do to help?



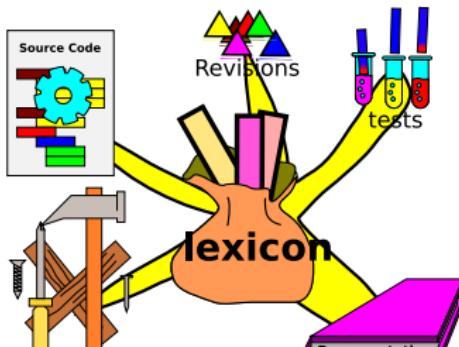
tagging
helps
RUPV



store your
artifacts
in accessible
repositories

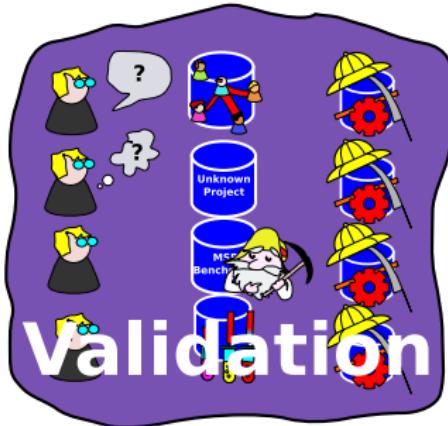


tools can
enable more
accurate
traceability links

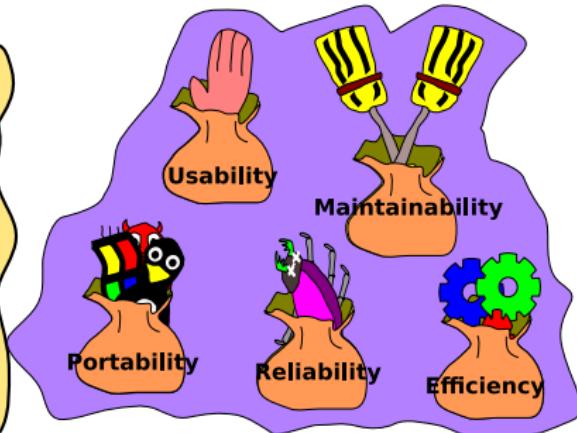
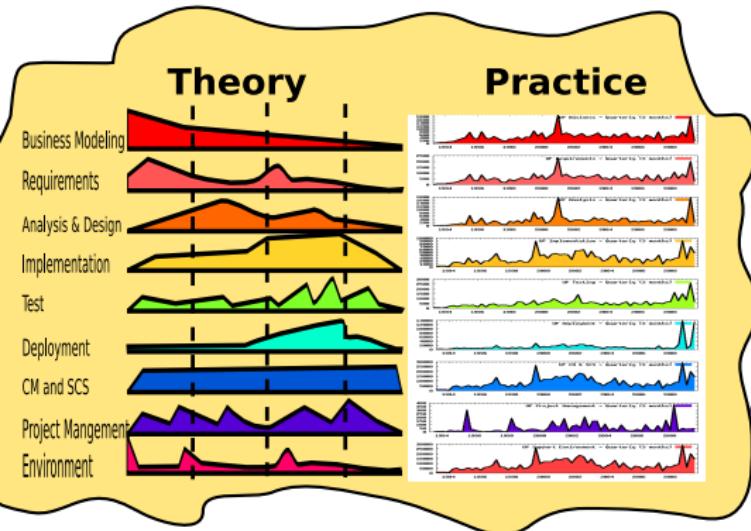


Tools can help enforce
a common lexicon
across projects.

Future Work



Conclusions



Related publications

- RUPV: Submitted to ICSM10
- NFR topic labels: Submitted to FSE10
- Developer Topics: ICSM09
- Release Patterns: MSR07 & ICSM07
- Maintenance Categories: ICPC09 & MSR08

